Proceedings of the

THIRD ANNUAL CONFERENCE OF THE COGNITIVE SCIENCE SOCIETY

BERKELEY, CALIFORNIA AUGUST 19-21, 1981

THE THIRD ANNUAL CONFERENCE OF THE COGNITIVE SCIENCE SOCIETY

Berkeley, California
August 19-21, 1981

The Conference is supported in part by a grant from The Alfred P. Sloan Foundation.

TABLE OF CONTENTS

MAJOR ADDRESSES

Constraint, Construal, and Cognitive Science 1
Robert P. Abelson

SYMPOSIUM — GOALS

A Model for Planning in Everyday Situations 11
Robert Wilensky

A Commitment-Based Framework for Describing Informal Cooperative Work 17
Richard E. Fikes

A Cognitive Science Approach to Improving Planning. 23
Barbara Hayes-Roth

Everyday Problem Solving. 29
James A. Levin

Marriage is a Do-It-Yourself Project: The Organization of Marital Goals 31
Naomi Quinn

SYMPOSIUM — COGNITION AND PERCEPTION

Transformational Structure and Perceptual Organization 41
Stephen E. Palmer

Rational Processes in Perception 50
Alan Gilchrist, Irvin Rock

The Role of Spatial Working Memory in Shape Perception 56
Geoffrey E. Hinton

Color Perception and the Meanings of Color Words. 61
Paul Kay

Structure and Function in the Early Processing of Visual Information 65
Shimon Ullman

SYMPOSIUM — AFFECT

Affect, Emotion, and Other Cognitive Curiosities. 75
George Mandler

Affect and Memory Representation 78
Wendy G. Lehnert

Situation Based Emotion Frames and the Cultural Construction of Emotions 84
Catherine Lutz

Disentangling the Affective Lexicon 90
Andrew Ortony, Gerald L. Clore

Affect and the Perception of Risk. 96
Amos Tversky, Eric J. Johnson

SYMPOSIUM — MENTAL MODELS OF PHYSICAL PHENOMENA

Generative Analogies as Mental Models 97
Dedre Gentner

The Role of Experience in Models of the Physical World 101
Andrea A. diSessa

The Form and Function of Mental Models 103
P. N. Johnson-Laird

Learning Through Growth of Skill in Mental Modeling 106
Jill H. Larkin, Herbert A. Simon

SUBMITTED PAPERS

Interestingness and Memory for Stories 113
John B. Black, Steven P. Shwartz, Wendy G. Lehnert

Use of Goal-Plan Knowledge in Understanding Stories 115
Edward E. Smith, Allan M. Collins

Inferences in Story Comprehension 117
Raymonde Guindon (sponsored by Alice Healy)

Self-Embedding is *Not* a Linguistic Issue 121
John M. Carroll

Center-Embedding Revisited 123
Michael B. Kac

The Comprehension of Focussed and Non-Focussed Pronouns 125
Jeanette K. Gundel, Deborah A. Dahl

The Natural Natural Language Understander 128
Henry Hamburger, Stephen Crain

Why Do Children Say "Goed"? — A Computer Model of Child Generation 131
Mallory Selfridge (sponsored by Katherine Nelson)

Five Experiments in the Development of the Early Infant Object Concept 133
George F. Luger, T. G. R. Brower, Jennifer G. Wishart

Analogy Generation in Scientific Problem Solving 137
John Clement

Understanding Design 141
Gerhard Fischer, Heinz-Dieter Boecker

System Properties of American Law: Constraints on Applying Cognitive Science 143
Janet L. Lachman, Roy Lachman

Writing with a Computer 145
Ira Goldstein

Creating Pleasant Programming Environments for Cognitive Science Students 148
M. Eisenstadt, J. H. Laubsch, J. H. Kahney

New Tools for Cognitive Science 150
Leonard Friedman

Cultural Constraints on Cognitive Representation 152
Alf Zimmer

Some Thoughts on Logic, Language, Mind and Reality 156
Rachel Joffe Falmagne

Time and Cognition: The Domestication of the Maya Mind 159
Hugh Gladwin

Why Some Modified Class Inclusion Tasks are Easy for Young Children:
A Process Model for Finding Referents of Labels in Arrays 163
Adele A. Abrahamsen

MOPs and Learning 166
Roger C. Schank

Retrieving General and Specific Information from Stored Knowledge of Specifics . . . 170
James L. McClelland

Can *If* be Formally Represented? 173
D. S. Brée

To See And Not to See, That is the Question 177
E. Andreewsky, A. Andreewsky, G. Deloche, D. Bourcier

E/Motional Memory as a Mediating Construct in the Study of Person/Environment
Interaction 179
Marisa Zavalloni

The Perception of Disoriented Complex Objects 181
Steven P. Shwartz (sponsored by Stephen Kosslyn)

The Role of Intrinsic Axes in Shape Recognition 184
Marianne Wiser (sponsored by Mary Potter)

Relations Between Schemata-Based Computational Vision and Aspects of Visual
Attention 187
Roger A. Browse (sponsored by D. Kahneman)

Growing Schemas Out of Interviews 191
Jerry R. Hobbs, Michael H. Agar

Shaping Explanations: Effects of Questioning on Text Interpretation 193
Richard H. Granger, Jr. (sponsored by Dogulas R. White)

The Need for Context in Event Identity 197
John M. Morris

Point of View in Problem Solving 200
Edwin L. Hutchins, James A. Levin

Representing Problem-Solving Episodes 203
Arthur M. Farley, David L. McCarty

Where Process- and Measurement Models Meet:
Evaluation of States in Problem Solving 207
Jan Drösler (sponsored by Walter Kintsch)

Positive Affect and Creative Problem Solving 210
Alice M. Isen, Gary P. Nowicki

Memory in Story Invention. 213
Natalie Dehn (sponsored by Robert Abelson)

Recognizing Thematic Units in Narratives. 216
Brian J. Reiser, Wendy G. Lehnert, John B. Black (sponsored by Andrew Ortony)

Using Qualitative Simulation to Generate Explanations 219
Kenneth Forbus, Albert Stevens

Incentive and Cognitive Processing 222
Michael W. Eysenck

The Role of TAUs in Narratives 225
Michael G. Dyer (sponsored by Wendy Lehnert)

Controlling Parsing by Passing Messages 228
Brian Phillips, James Hendler

A Parser with Something for Everyone 231
Eugene Charniak

Thought Sequences and the Language of Consciousness 234
Benny Shannon

What Conditions Should a Theory of Consciousness meet?. 236
Bernard J. Baars

The Role of the Observer in Cognitive Science 238
Richard Hammersley (sponsored by Donald Norman)

Are We Ready for a Cognitive Engineering? 240
S. K. Card, A. Newell

Knowledge Structures of Computer Programmers. 243
Beth Adelson (sponsored by Marilyn Shatz)

GLISP: An Efficient, English-Like Programming Language. 249
Gordon S. Novak, Jr.

Some Questions About Verbal and Pictorial Representation. 252
Sheldon Richmond

The Relationship Between Human Motion and Objects in the Cognitive
Representation of Visual Events 255
Margot D. Lasher

Toward a Model of Cognitive Process in Cartoon Comprehension. 261
Michael E. J. Masson, Inga Boehler (sponsored by Peter Polson)

Demon Timeouts
Limiting the Life Span of Spontaneous Computations in Cognitive Models 263
Steven Small (sponsored by Chuck Rieger)

The Effects of Integrated Knowledge on Fact Retrieval and Consistency Judgments: When Does it Help, and When Does it Hurt? 265
Lynne M. Reder, Brian H. Ross

A Question Answering Method of Exploring Prose Comprehension: An Overview ... 268
Arthur C. Graesser

Role of Context in Cognitive Development 270
Gideon Carmi

Concept Formation Through the Interaction of Multiple Models 271
Mark H. Burstein (sponsored by Christopher Riesbeck)

A Simulated World for Modeling Learning and Development 274
Pat Langley, David Nicholas, David Klahr, Greg Hood

Why Do We Do What We Do? 277
James A. Galambos, John B. Black (sponsored by Lance Rips)

An Activation-Trigger-Schema Model for the Simulation of Skilled Typing 281
David E. Rumelhart, Donald A. Norman

Toward a Formal Theory of Human Plausible Reasoning 283
Allan Collins, Ryszard S. Michalski

Skills, Learning and Parallelism 284
Aaron Sloman

Cognitive Style and the Learning of a First Computer Language 286
M. J. Coombs (sponsored by R. Wilensky)

The Concept of E-Machine: On Brain Hardware and the Algorithms of Thinking 289
Victor Eliashberg

Invariance Hierarchies in Metaphor Interpretation 292
Jaime G. Carbonell (sponsored by Philip Hayes)

Dual-System Processing in Language Comprehension 296
Jon M. Slack

Metaphor as Nonliteral Similarity 299
Odella E. Schattin, Hartvig Dahl (sponsored by Virginia Teller)

A Theory of Intelligence 301
P. J. vanHeerden

A Tetragenic Frame for Modelling Unmediated Knowledge Acquisition 303
Egon E. Loebner

Surrealistic Imagery and the Calculation of Behavior 307
Sheldon Klein, David A. Ross, Mark S. Manasse,
Johanna Danos, Mark S. Bickford, Walter A. Burt, Kendall L. Jensen

Learning with Understanding 310
James G. Greeno

The Growth of Number Representation: Successive Levels of Schematic Learning. . . 313
Lauren B. Resnick

The Role of Experiences and Examples in Learning Systems 317
Edwina L. Rissland, Oliver G. Selfridge, Elliot M. Soloway

I-Interruption Effects in Backward Pattern Masking:
The Neglected Role of Fixation Stimuli 320
J. Pascual-Leone, J. Johnson, D. Goodman, D. Hameluck, L. H. Theodor

Cognitive Load Affects Early Brain Potential Indicators of Perceptual Processing. . . . 323
Francois Richer, Jackson Beatty

The Spatial Representation and Processing of Information in Cognition. 326
Gary W. Strong, Bruce A. Whitehead (sponsored by John Holland)

MAJOR ADDRESSES

CONSTRAINT, CONSTRUAL, AND COGNITIVE SCIENCE

Robert P. Abelson
Yale University

Cognitive science has barely emerged as a discipline -- or an interdiscipline, or whatever it is -- and already it is having an identity crisis.

Within us and among us we have many competing identities. Two particular prototypic identities cause a very serious clash, and I would like to explicate this conflict and then explore some areas in which a fusion of identities seems possible. Consider the two-word name "cognitive science". It represents a hybridization of two different impulses. On the one hand, we want to study human and artificial cognition, the structure of mental representatives, the nature of mind. On the other hand, we want to be scientific, be principled, be exact. These two impulses are not necessarily incompatible, but given free rein they can develop what seems to be a diametric opposition.

The study of the knowledge in a mental system tends toward both naturalism and phenomenology. The mind needs to represent what is out there in the real world, and it needs to manipulate it for particular purposes. But the world is messy, and purposes are manifold. Models of mind, therefore, can become garrulous and intractable as they become more and more realistic. If one's emphasis is on science more than on cognition, however, the canons of hard science dictate a strategy of the isolation of idealized subsystems which can be modeled with elegant productive formalisms. Clarity and precision are highly prized, even at the expense of common sense realism. To caricature this tendency with a phrase from John Tukey (1969), the motto of the narrow hard scientist is, "Be exactly wrong, rather than approximately right".

The one tendency points inside the mind, to see what might be there. The other points outside the mind, to some formal system which can be logically manipulated (Kintsch et al., 1981). Neither camp grants the other a legitimate claim on cognitive science. One side says, "What you're doing may seem to be science, but it's got nothing to do with cognition." The other side says, "What you're doing may seem to be about cognition, but it's got nothing to do with science."

Superficially, it may seem that the trouble arises primarily because of the two-headed name cognitive science. I well remember the discussions of possible names, even though I never liked "cognitive science", the alternatives were worse; abominations like "epistology" or "representonomy".

But in any case, the conflict goes far deeper than the name itself. Indeed, the stylistic division is the same polarization than arises in all fields of science, as well as in art, in politics, in religion, in child rearing -- and in all spheres of human endeavor. Psychologist Silvan Tomkins (1965) characterizes this overriding conflict as that between characterologically left-wing and right-wing world views. The left-wing personality finds the sources of value and truth to lie within individuals, whose reactions to the world define what is important. The right-wing personality asserts that all human behavior is to be understood and judged according to rules or norms which exist independent of human reaction. A similar distinction has been made by an unnamed but easily guessable colleague of mine, who claims that the major clashes in human affairs are between the "neats" and the "scruffies". The primary concern of the neat is that things should be orderly and predictable while the scruffy seeks the rough-and-tumble of life as it comes.

I am exaggerating slightly, but only slightly, in saying that the major disagreements within cognitive science are instantiations of a ubiquitous division between neat right-wing analysis and scruffy left-wing ideation. In truth there are some signs of an attempt to fuse or to compromise these two tendencies. Indeed, one could view the success of cognitive science as primarily dependent not upon the cooperation of linguistics,AI, psychology, etc., but rather, upon the union of clashing world views about the fundamental nature of mentation. Hopefully, we can be open minded and realistic about the important contents of thought at the same time we are principled, even elegant, in our characterizations of the forms of thought.

The fusion task is not easy. It is hard to neaten up a scruffy or scruffy up a neat. It is difficult to formalize aspects of human thought which are variable, disorderly, and seemingly irrational, or to build tightly principled models of realistic language processing in messy natural domains. Writings about cognitive science are beginning to show a recognition of the need for world-view unification, but the signs of strain are clear. Consider the following passage from a recent article by Frank Keil (1981) in Pscyhological Review, giving background for a discussion of his formalistic analysis of the concept of constraint:

"Constraints will be defined...as formal restrictions that limit the class of logically possible knowledge structures that can normally be used in a given cognitive domain." (p. 198).

Now, what is the word "normally" doing in a statement about logical possibility? Does it mean that something which is logically impossible can be used if conditions are not normal? This seems to require a cognitive hyperspace where the impossible is possible.

It is not my intention to disparage an author on the basis of a single statement infelicitously put. I think he was genuinely trying to come to grips with the reality that there is some boundary somewhere to the penetration of his formal constraint analysis into the viscissitudes of human affairs. But I use the example as symptomatic of one kind of approach to the cognitive science fusion problem: you start from a neat, right-wing point of view, but acknowledge some limited role for scruffy, left-wing orientations. The other type of approach is the obvious mirror: you start from the disorderly left-wing side and struggle to be neater about what you are doing. I prefer the latter approach to the former. I will tell you why, and then lay out the beginnings of such an approach.

The strategy of trying to move leftward from the right suffers from a seemingly permanent limitation on the kinds of content and process you are

willing to consider. If what really matters to you is the formal tractability of the domain of investigation, then your steps are likely to be small and timid. Recent history in several social and behavioral science areas makes this quite clear.

In cognitive anthropology, there was a great deal of fascination 25 years ago with the orderliness of systems for kinship terminology. Kin terms in different societies were found to be precisely describable by concatenations of the values on a handful of well-specified components such as sex and generation. Formal mini-models captured these regularities elegantly. Originally it was thought that this kind of componential semantics held great promise for the analysis of culture and language in general, but gradually it was realized that outside of kinship terms and pronoun systems, precious little else in the language of any society was ordered so neatly. Faith in tight componential analysis has largely been abandoned.

Rational decision theory has until recently had a tight hold on the views of economists and some psychologists of the way people made decisions. The typical rational decision model specifies a set of uncertain outcomes, with each of which is associated a probability and a utility. Choices among ensembles ot outcomes are then said to be predictable on the basis of a strict composition rule on the probabilities and utilities. The only trouble is, the behavior of human subjects overwhelmingly disobeys the predictions of the models, no matter how hard the axioms try. There have been numerous attempts to rescue the general framework, including the clever strategy of training subjects to obey the rational model following initial demonstrations of deviation from it. I would recommend this device also to people promoting competence models of syntax in the face of incompetent performers, except that I cannot as a psychologist bring myself to believe that it tells us anything about human psychology. In any case, there are some many violations of rational decision theory that it is a clear failure as a descriptive or explanatory psychological model. Only an approach that deals directly with observed decision phenomena (for example, the work of Tversky and Kahneman (1980)) has a chance of success. (For fuller reviews of this field, see Einhorn and Hogarth (1981), March (1978), and Abelson & Levi (Note 1)).

Other examples of excessive faith in the unaided power of formalisms to subdue the beast of psychological explanation could be adduced from within experimental psychology itself. A good case from some years back is provided by stochastic learning models (Bush & Mosteller, 1955), which were extremely rich as mathematical objects,but turned out to have applicability to a very small range of problems, indeed. Models like this were part of the "bottom up" tradition of doing science within experimental psychology, the belief that by starting with very tightly controlled, limited, and isolated laboratory phenomena, one could gradually explicate the operation of the whole organism. This tradition is of course still strongly honored by many experimental pscyhologists, but I think that those psychologists interested in cognitive science have largely departed from that tradition, at least in its most extreme form. In the service of studying more important and more general phenomena than those falling within the formal boundaries of mini-models, cognitive science psychologists have been willing to use messier stimulus materials and at least contemplate non-laboratory methodologies. The way is not easy, and there is much anguishing. That, I claim, is the price of trying to move leftward from a right-wing starting point.

Linguists, by and large, are farther away from a cognitive science fusion than are cognitive psychologists. The belief that formal semantic analysis will prove central to the study of human cognition suffers from the touching self-delusion that what is elegant must perforce be true and general. Intense study of quantification and truth conditions because they provide a convenient intersection of logic and language will not prove any more generally informative about the huge range of potential uses of language than the anthropological analysis of kinship terms told us about culture and language. On top of that, there is the highly restrictive tradition of defining the user of language as a redundant if not defective transducer of the information to be found in the linguistic corpus itself. There is no room in this tradition for the human as inventor and changer and social transmitter of linguistic forms, and of contents to which those forms refer. To try to understand cognition by a formal analysis of language seems to me like trying to understand baseball by an analysis of the physics of what happens when an idealized bat strikes an idealized baseball. One might learn a lot about possible trajectories of the ball, but there is no way in the world one could ever understand what is meant by a double play or a run or an inning, much less the concept of winning the World Series. These are human rule systems invented on top of the physical possibilities of the batted ball, just as there are human rule systems invented on top of the structural possibilities of linguistic forms. One can never infer the rule systems from a study of the forms alone.

Well, not I have stated a strong preference against trying to move leftward from the right. What about the other? What are the difficulties in starting from the scruffy side and moving toward the neat? The obvious advantage is that one has the option of letting the problem area itself, rather than the available methodology, guide us about what is important. The obstacle, of course, is that we may not know how to attack the important problems. More likely, we may think we know how to proceed, but other people may find our methods sloppy. We may have to face accusations of being ad hoc, and scientifically unprincipled, and other awful things.

Sometimes we worry about such matters ourselves. There is a neat person struggling to get into every scruffy person (just as there is a scruffy person struggling to get out of every neat person). What is required is that we act on our worries, that we try to take the criticisms seriously and see what can be done about them. The messy intuitions and theories, albeit they concern very general and important problems (God bless 'em), need to be articulatued and developed in a more orderly way.

I will take the work of the Yale Artificial Intelligence Project, and an particular, the programmatic statements in the Schank and Abelson (1977) book as point of departure. The Yale point

of view is quintessentially scruffy, and has been criticized accordingly. No matter that scrips and plans and goals and themese are psychologically reasonable, and that computer programs using such concepts are operational at the frontier of realistic processing of natural language, nevertheless, it is said that the system of concepts is not formal enough.

The make a system more formal is to define its concepts more precisely, and to have them enter into general predictions and explanations according to a set of principles, preferable a small elegant set. Let me first address the question of the definition of concepts. The Yale group deals with knowledge structures such as scripts and plans, it may seem at first that these are pretty amorphous entities. What counts, say, as a script or as clearly not a script? How can you tell? When we gave examples such as the restaurant script, and cognitive psychological experiments (e.g., Bower, Black & Turner, 1979; Galambos & Rips, 1979; Graesser, 1981) forthwith used verbal stimuli from the restaurant situation, along with doctor visits and laundromat activities and so on, it may have produced misleading impressions that the intention was to define scripts solely by waving at passing examples, or perhaps by writing down a definitive list of lll of them, or worse, by allowing any damn thing to be a script just by calling it that.

I say that these impressions are misleading because in fact we have become acutely aware (Schank, Note 2) that scripts have been loosely used. The original intention was not at all to create a haven for loose concepts; in fact, scripts (among other knowledge structures) were very tightly defined, by a set of interdependent constraints. Indeed, if a knowledge structure is proposed as crucial in the top-down processing of certain inputs, then clearly it must embody of us to leave these constraints largely implicit, rather than spelling them out systematically.

It is not my main intention today to remedy this neglect for scripts or any other specific type of knowledge structure, but rather to make clear my general view of the role that constraints play in the process of understanding text. However, having raised the issue, it is useful to begin by indicating what constrains script structures. Related remarks apply to related types of structures such as MOPs (Schank, 1980) and metascripts (Abelson, 1981).

The casual definition of a script is "a stereotyped sequence of events familiar to the individual". Implicit in this definition are two powerful sources of constraint. One is the notion of an event sequence, which implies the causal chaining of enablements and results for physical events and of initiations and reasons for mental events. Causal chains are highly ruleful, and many of those rules have been spelled out explicitly (Schank, 1975; Schank and Abelson, 1977, Ch.2). The other constraint generator comes from the ideas of stereotypy and familiarity. That an event sequence is stereotyped implies the absence of fortuitous events. Also, for events to be often repeated implies that there is some set of standard individual and institutional goals which gives rise to the repetition. Furthermore, there are almost certainly subgoals, each of which defines a scene involving a transaction between particular role players in a certain physical setting, using given props.

At the scene level of a script, therefore, there run in parallel four networks of coherences: Those governing the transactions, the role players, the physical settings, and the props. None of these entities can enter into sequences arbitrarily. Scene transitions between one physical setting and another, for example, follow the topographical rules of familiar environments. One does not step off the airplane directly into a swimming pool, or go from the doctor's waiting room into the kitchen. Role players remain from scene to scene except when somebody makes a purposive and expected entry or exit. It does not require belaboring these coherences in full detail to realize the enormous degree of constraint thus imposed on input relevant to any given script.

Perhaps on of the things that disguised the high order of systematicity of scripts was that some of the computer programs that used them, such as SAM (Cullingford, Note 3), were written in a way that did not insure against ad hoc violations of some of the constraints. A prankish programmer could perfectly well prepare an expected event sequence wherein the customer ate the check and gave the food to the cashier, thus thwarting their mutual goals, and nothing in the Script Applier would protest illegitimacy of such expectations. Of course they would turn out by experience to be useless in matching realistic inputs, but that is a very weak way to recognize absurdity. (And it is still vulnerable to the possibility of prankish inputs). Later programs such as POLITICS (Carbonell, 1978) did not, by the way, suffer the same degree of vulnerability, but this whole issue has not been treated as explicitly as it might be.

Why is any of this important? Well, there may be some people who feel it is not important, that there are more compelling issues for language AI to worry about. But it bothers me that the concept of structural constraint seems to have been coopted by the neats, when all the while the scruffy Yale programs are based very heavily on a whole series of implicit constraints.

Let us look more closely at some general issues pertaining to the idea of constraint. In the original formulation of information theory and communication theory, the structural constraints on the communicative elements were presumed mutually accessible to the sender and receiver of messages. They each knew the redundancies of letter strings or phoneme strings, and this consensus was the basis for an analysis of the information content of messages. In effect, one could ignore most of the properties of the receiver, and concentrate the analysis on the properties of the stimulus ensemble.

Nowadays the emphasis in cognitive science is on chunks of meaning, and one cannot generate meaning simply by higher-order approximations to the structure of low-level stimulus elements. The idea that the set of possible messages is very much constrained is still a powerful idea, but at least two drastic changes are necessary in applying an information theoretic type of analysis to higher-level meaning elements, say,

sentences, rather than to low-level stimulus elements such as letters. For one thing, the number of possible elements is infinite rather than finite. For another, there is no guarantee at all that the receiver of messages adequately comprehends the structure of contingencies between sentences that can possible be generated by the message source.

People in the Chomskyan tradition writing about constraints in knowledge structures do not usually distinguish between structural constraint intrinsic to the stimulus ensemble and structural contraint characterizing the receiver's construal of the stimulus possibilities. The former focuses the analysis structure strictly on analysis of language, completely defining the psychology of the receiver out of independent existence. This is a very right-wing attitude, in the sense I have previously discussed. People are said to be the way they are because of immutable external regularities. There is little interest in studying learning, or human error, or individual differences in intelligence.

Furthermore, there is total disregard of cultural shaping of knowledge structures. That is, even in cases where there is a structural match between the semantics of the language and the corresponding mental representation in a particular domain, this match may have been produced by a process of cultural invention rather than by the inevitable emergence of a natural truth. Much social knowledge pertains to what anthropologists (cf. D'Andrade, Note 4) call constitutive rule systems, extensive networks of how to construe, how to behave in, and even how to feel about culturally defined situations. The nexus of rules defining the meaing of marriage is one example. Other examples of cultural rule systems are mental illness, senior citizenship, golf, and sexual harassment. It seems to me perfectly obvious that there is no foreordained meaning for any of these domains, or a thousand others, rather, a meaning which evolves under the pressure of social, political and economic motives and experiences. I belabor the banal here because of recently renewed claims that to know knowledge, one only need know semantics.

Having thus gored the one-horned ox, let me try to lay out a balanced view of one aspect of the interplay between mental representatives and stimulus structures. I will place the argument in the context of text understanding.

Consider an individual who is exposed to a string of language, presented one chunk at a time, say, sentence by sentence. Considerable constraint will be imposed by the general context surrounding the presentation of the string, say, whether it is a story or a piece of conversation, and what nature of the topic and style of presentation. The local context operative at a particular place within the string will exercise further constraint. Is there a way to conceptualize a measure of the degree of structuredness at any given place in the presentation of the string? Further, is there a way to think about structure such that it is a joint property of the stimulus string and the interpretive machinery of the understander?

I propose a characterization relating the structuredness of a context to the constraint in the stimulus string and something I will call the construal function on the part of the understander. The constraint in the stimulus string can be expressed by the distribution of probabilities P(i) of occurrence over all potential next stimulus chunks. If a few inputs are moderately likely and all others are of very low probability, the stimulus contains more structure than if likelihood is fairly evenly distributed over a large set of possibilities. This is as in standard information theory.

But the understander may not extract from the stimulus the available structure. The individual has expectations of what sorts of things may occur next. If something which is highly expected occurs, it is difficult to process. In general, we may imagine that there is a distribution of measures of processing facility F(i) over all possible next stimulus chunks. Under various different construals of what the stimulus string is about, thus what expectations are appropriate, the distribution of facility measures will be different. Perhaps processing facility could be operationalized as the speed with which a given chunk can be comprehended, or perhaps in some more subtle way, but in any case the understander is conceptualized as having the capacity to prepare for coming inputs by making a differential allocation of facility measures F(i). This is a much more realistic view of the understander than assuming that he has no expectations at all, thus relatively equal facility in accepting all inputs, or only a single dominant expectation. Artificial intelligence programs that work heavily top-down always in effect smear their expectations over a domain of related possibilities. A good image for this emaring tendency is Chuck Reiger's concept of the "expectancy cloud".

The average value of processing facility is the sum of cross-products of stimulus probability P(i) and facility measure F(i) over all possible input chunks. This average facility will be high or low depending on four things: (1) the general simplicity of the context; (2) the general facility of the understander; (these two factors can jointly be characterized by the mean of the F's unweighted by stimulus probability): (3) the degree of predictability inherent in the stimulus string (which could be indexed by an uncertainty measure on the P's); and (4) the match between the F's and the P's that is, the extent to which the construal by the understander appropriately allocates her preparations in the direction of inputs which are relatively likely to occur. Under an assumption of a fixed unweighted variance of the F's, it is easy to show that the average proportionality relationship between the P's and the F's, high facility attaching to relatively high likelihood, and low facility attaching to relatively low likelihood.

The match between what is expected and what might occur should not automatically be presumed, either according to some cosmic principle of innate resonance between the individual and the environment, or on the basis of a gradual learning of, and accomodation to contextual contingencies. There are at least four reasons for this. First, it is very possible, indeed frequent, for people to misconstrue situations and have a whole series of misguided expectations.

Misconstrual tendencies are very interesting to social psychologists and there has been a good deal of recent research on stereotyping, on misleading first impressions, on the effects of inappropriate but salient schemata, and on the insensitivity of false constructions to empirical evidence (cf. Nisbert & Ross, 1980).

A second reason not to presume that construals reflect stimulus constraints is that people are generally extremely slow to pick up the contingency structure in novel input materials, if they ever get it at all. Contingencies are especially problematic when multidimensional dependencies are involved. It is clear from classical two-alternative guessing situations that people are very good at learning the simplest zero-order structure, that is, the relative frequencies of two different alternatives. Although appropriate data do not to my knowledge exist, there is little reason, however, to suppose that people are adept at learning the zero-order structure within large numbers of alternatives. And it is very clear from so-called "cue validity" studies (e.g.,Hammond & Summers, 1965) that there are sharp limitations on the learning of higher-order structures. When many independent cues are modest predictors of an outcome variable, people are unable to use all the cues, but settle instead for (somewhat fallible) use of three or four of them. In realistic stimulus domains where it is not at all clear how many cues of what sort there might be, the situation can be even worse. For example, in studies of how people judge whether someone else is lying or not (Krauss et al., 1976; Kraut, 1978), the facial, gestural and speech cues that judges employ to diagnose the liar overlap hardly at all with the set of cues that actually predict lying.

A third reason, related to the second, is that stimulus structure is usually dependent on the source of the string being communicated. Different communicators have different styles and different angles on what to say or write about a given topic. The understander usually will not have long enough experience with particular communicators to pick up their individual contingency structures, even if learning were very rapid.

Fourth, the construal function must be flexible in operation, so that when there is a shift in the topic of the stimulus string, the understander can establish a new set of expectations. A lack of matching could come about if adjustments in construal were sluggish, lagging behind the stimulus. It is intuitively clear that there are both individual and situational differences in the rate of adjustment of construals. Part of the ordinary concept of intelligence, or perhaps quick-mindedness, is the ability ot an individual to keep up with what is being read or said, especially if the point is rapidly shifting. Situations, too, may help or hinder quick reconstrual. There is a good deal of psychological literature on the phenomenon of perserverance, wherein a person's analysis of a problem area continues in a vein which has previously been successful, despite the introduction of new materials which make the old analysis outmoded (Luchins, 1942), or the presentation of evidence that past success was spurious (Ross, Lepper, & Hubbard, 1975).

In short, the structure in personal construals need not match the structure of stimulus constraints, for several reasons. When there is a match, however, understanding is considerable facilitated. The example of script processing provides an especially clear case. An account of a highly scripted activity such as a visit to a doctor introduces very high stimulus constraint, because only a limited number of events have high probability of occurrence in the account. If the understander construes the account as indeed concerning a doctor visit (as opposed, say, to a chat with a professional colleague), then his relevant knowledge structure will highly constrain his expectations to a small set of events. Given a sufficient consensus on what sorts of things transpire in accounts of doctor visits, understanding will (on average) be very facile.

My discussion to this point has concealed an important aspect of the concepts of constraint and construal in text processing. There are really three different types of structural limitation on coherent stimuli and coherent expectations. Recall that we are supposing that the input string is received a chunk at a time, and that we are interested in the probability of occurrence and the processing facility associated with every possible chunk. For tangibility we may suppose that the chunks are sentences. Two somewhat different kinds of constraints are those applying within chunks, and those applying between chunks. Let us designate these respectively as combinatorial constraint and sequential constraint. A third kind I will call functional equivalence constraint, to which I will return shortly.

By within chunk, or combinatorial constraint, I refer to tendencies or rules for what linguistic components go with what. This would include all of syntax, semantic rules or "selection restrictions" about sensible meaning combinations, such as what actions require animate actors and what attributes are pertinent for what object classes, and also fragments of pragmatic knowledge that tell us what combinations are unlikely in the real world even though semantically possible, such as the Queen using obscenity or coal miners curtsying. In this category of constrint, it seems quite likely that rules characterizing stimulus structure would be generally well matched by rules characterizing construals. These aspects of constraint are widely appreciated and shared, and the reasons given above in support of mismatch tendencies are least likely to apply.

By between chunk, or sequential constraint, I mean tendencies for certain chunks to follow a given chunk or sequence of chunks. From a formal point of view, one might suppose that between chunks constraint is just another bundle of selection restrictions, or set of rules about what goes with what, and thus is just like within chunk constraint but operating on larger units. In the field of story understanding, the supposition has sometimes been made that a kind of syntax exists linking successive units, and so we have the notion of "story grammars" (cf. Rumelhart, 1977; Mandler & Johnson, 1977).

Story grammars may have some use as rough tools in restricted story domains, but they have I think rightly been criticized (e.g., Black & Wilensky, 1979) when they make overly strong claims. Stories are simply not grammatical in quite the same sense that sentences are. Nor are other bodies of text.

When a literate reader asserts that a short string of writing within a period at the end is not a sentence, the literate observe will almost always agree. But who is to say that a longer string of writing is not a paragraph, or not a story. Judgements here are somewhat fuzzy, as we are dealing not with inviolate rules but with general coherence tendencies. Pragmatics and stylistics become much more important, and the role of syntax dwindles. Construals can more readily be out of synch with stimulus constraints. To think otherwise,and to regard sequential constraint as merely an extension of combinatorial constraint, is to fall too much under the thrall of the bottom-up, start small, scientific strategy of the neats.

The third category, that I have called "functional equivalence" constraint, is rather different in nature. In the general conception I have sketched of the construal process, there is the clear unrealism of supposing that the organism allocates preparatory provisions among an infinite number of possible stimuli. Much more reasonable is the supposition that the understander prepares differentially for different categories of input content, in effect grouping stimuli into equivalence classes according to the functions they serve. Within each class, processing facility would be roughyl constraint. This is a fairly strong form of subjectively imposed constraint, for it says that such-and-such stimuli are to be regarded as equivalent, and all processed with ease, whereas such-and-such other stimuli, constituting another equivalence class, will be processed with less facility, and so on. Now there is great opportunity for mismatching, depending on whether the understander does or does not carve the possibilities at different joints than the probability structure of the stimuli would. To maintain matching, it would be necessary for every stimulus within an equivalence class to be approximately equally probable. The understander, however, may lump possibilities together because they have comparable personal concern or affective significance, not because they are necessarily objectively substitutable.

These ideas may become clearer if I recapitulate and then amplify the model I am outlining. We are supposing an understander exposed to a stimulus string one chunk at a time. Strong structural constraints characterize the ensemble of stimulus strings which the given string instantiates, but we do not presuppose that the understander necessarily perfectly appreciates those constraints. Instead, the understander, in a construal process, or general expectational policy, imposes constraints of her own. Construals are from time to time altered as the stimulus string unfolds, but while in force, each given construal is a processing allocation which determines the particular degree of facility with which different possible inputs and sequences of inputs would be processed. A construal defines functional equivalence classes of stimuli, such that stimulus possibilities within class are processed alike.

The average facility of processing is maximized when there is a match between the probability structure of the stimulus ensemble and the facility structure of the construal. Furthermore, given some degree of match, the average facility increases with increasing structuredness. There are really three aspects of both types of structure: the zero-order structure partitioning the possibilities according to how likely or expected they are, and the contingency structure limiting what goes with what, both within chunks and across chunks.

I have argued that matching between construal structure and stimulus constraint structure should not in general be presupposed, but that it is likely to occur in certain contexts. Within chunks, it is likely that syntactic constraints will be well matched. Across chunks, certain types of knowledge structures permit ready construals of well-structured realities. I have already mentioned scripts as one example.

Another between-chunk example has to do with the role that knowledge of intentionality plays in story understanding. Stories tend to be summarized in terms of goals and plans of the main characters. Goals and successful plans tend to be well remembered. AI models of story understanding make a big point of tracking goals and goal fates. This was the case in Schank and Abelson (1977), and in other Yale models such as those of Wilensky (1978(and Carbonell (1978), and in a great many non-Yale models as well. Why is this? What is there about goals that makes them so important? From a scruffy common sense point of view, this question may not seem worth asking because the answer seems obvious. Goals underly most of the activities of people, and what interests people is hearing about things that other people do. One might amplify this intuition with the observation that intentionality is not observable, but must be inferred, and there is something especially intriguing in making inferences about people to explain why they do what they do. Another angle is that goals relate to emotions, as I will discuss later, and emotions are especially interesting.

A formalist would not be especially happy with such explanations. They essentially say that goals are interesting because everybody knows they are interesting; but no principled account is given. Well, it seems to me clear that a principled account can be given, and indeed, it is implicit in almost all analyses of the role of goals and plans, but it is usually not spelled out. Simply put, it is that intentionality is "inference rich"--it provides a high degree of stimulus constraint. Construals structured to match will confer great advantage to the understander. Intentional action introduces constraints of all three kinds: combinatorial, sequential and functional equivalence. Especially noteworthy is the possibility of remote sequential constraint, that is, the influence of somebody having a goal on his actions much later in time. In a story or a novel-- or in life-- it could be dozens or hundreds of pages-- or days-- before the major goals of individuals are actualized, yet that latent potentiality is present all the while. This constraint demon that goals unleash is a kind of inferential time bomb set to go off one knows not when. No Markov process behaves like this.

It is clear, therefore, that knowledge structures concerning goals are highly constraining, thus very important in the understanding process. Whether goals are the most constraining concept in texts about human activities is very hard to say, because we have no comparative constraint measures on different inference-rich concepts, averaged over all possible, or all available, or all experienced texts of a given type. I want to point out, however, that a good

guiding philosophy behind the choice of knowledge structures to investigate is to try to pick those which are both highly constraining and highly frequent, thereby being very useful for construal functions. Although we did not articulate it explicitly, this was indeed the rationale guiding the Schank & Abelson (1977) choice of scripts, plans, goals, and themes as the highest priority knowledge structure concepts to investigage. Far from being ad hoc, therefore, these concepts are very closely tied to ideas concerning constraint.

Thus far in my account, I have given mainly a way to talk carefully about the role of structural constraint in the process of understanding, with very little in the way of predictive principles or substantive claims. Let me now attempt some claims based on the construal concept.

I have said that construals are from time to time altered or updated during the course of understanding. I believe that what is remembered about a stimulus string is not the string itself, but the set of construals and reconstruals used in interpreting it. When a construal is active and inputs arrive which are not readily processed, that is, are unexpected in terms of the construal, then reconstrual becomes necessary. Memory for one's own construal structures, therefore, would include both the original construal and the reconstrual compelled by unforeseen stimulus events. Thus my proposal here is similar to Schank & Abelson's (1977) script pointer plus "weird list" idea, and to Graesser's (1981) "script pointer plus tag" hypothesis, except that it is more general in that the knowledge structure involved need not be a script, but could be anything. Indeed, it could be any combination of knowledge structures implicated in a construal which partially succeeds and then needs correction. (Among other recent treatments containing similar ideas are those of Lehnert (1979) and of Schank (Note 2)).

There are some memory data which fit nicely with this Construal Principle (e.g., Hastile, 1980; Spiro & Eposito, 1977; Graesser, 1981). But I am also aware that there are other data which are hard to explain by it (e.g., Anderson & Pichert, 1978), and that there are complications in applying it to data in general. Let that be a story for another occasion, however.

Reconstrual seems to me to be an especially interesting phenomenon. One class I would like to pursue arises when two incompatible construals compete with one another for processing dominance. By incompatible construals, I mean sets of expectations based on the other set. The more massive the preparations, the more serious the incompatibility. In cognitive terms, the massive preparatory part of a construal is the establishment of sequential constraints; thus incompatible construals involve opposing sequential constraints.

Characteristically associated with incompatible construals, under certain conditions, are affective experience. There are different types of incompatibility, as I shall spell out, and with each of these is associated a different variety of affect. These connections are compelling enough to serve as the basis for a theory of affect. Recently, there have arisen in psychology a number of systematic taxonomies of affective states (deRivera, 1977; Roseman, 1979; Wilensky, Ortony & Collins, 1980) which set forth a number of disposing factors said to generate one or another affect. While highly evocative, the various schemes seem to lack a unifying principle common to their sets of affects, this I believe to be provided by the idea of incompatible construals.

My analysis owes much to the scheme devised by Roseman (1979), but is differently organized in order to get the benefit of the incompatibility principle. In conveying a preliminary version of my theory, I will continue to talk about construals, but I will depart freely from the text understanding pradigm to refer also to a behavioral situation paradigm where the prospective and actual events might happen or do happen to the individual involved, rather than his just imagining their occurrence for the characters in a story.

To avoid confuction, I should say what the theory is not about. It is not primarily about the pleasures and pains associated directly with physical sensations, either innately or thorough conditioning. Thus it is not primarily about sexual pleasure, or pleasurable tastes, or about fright, pain or disgust, or about the love or aversion for the people or objects associated with those pleasures or pains. Secondarily, however, it may implicate pleasures or pains or other goal states, as will be explained. The theory is also not about the semantics of affect as coded in words or phrases capable of evoking associated emotions, such a pejorative adjectives applied to another person so as to stimulate dislike or the sunny lyrics of a love song. Rather, it concerns the emergence of an affecgive state as a consequence of the structural relationships in an ongoing situation: it is a structural theory of "on-line affect."

In analyzing incompatible construals, we have to ask why more than one construal would ever be necessary. An obvious answer is that the ongoing construal leads to poor understanding, and must be changed. A more direct way to say this is that anticipation does not correspond to reality. What you imagine will happen does not in fact happen, and you must update your imaginings. If the update is incompatible with the previous construal, then an affective process will occur which is both a signal and a symptom of the activity of reconstrual (and which, incidentally, will be associated with high memorability for the event which precipitated the reconstrual).

Two conditions seem basic to the degree of incompatibility of construals. One is the range of possible cognitive chunks implicated by the two construals, the second is the discrepancy in the hedonic import of the two construals, whereby one is highly pleasurable or painful and the other is not. Inference-richness and hedonic import would seem in practice to cooccur, because one mainly makes extensive inferences about that which is personally consequential. But the two concepts are conceptually separable.

In any case, not every reconstrual involves compatibility, and many incompatibilities are quite trivial in extent and significance. Thus structural affect is not freely evoked by minor alterations from previous expectations. Thus if you mistook someone to say they were from Stanford

and it turned out they meant Stamford, (Connecticut), or if you thought that a session of the conference was in Room A and it turned out to be in Room B, those changes would not provoke affect (unless your mistake led to come commitment or consequence difficulty to undo).

It is instructive to consider systematically how the inference-richness of consequential alternative construals might vary and give rise to differing affective states. A useful rubric is the intentional action sequence, where a positive or negative outcome state is cognized by the individual along with an alternative outcome of opposite import. What is of interest is the point in the sequence at which the alternativity arises, thereby determining the depth of reconstrual which is necessary. There are different classes of cases, depending on whether two alternative construals are present only in imagination, or whether one is imagined and the other represents reality, or there are representations of two disparate realities because reality has changed.

Let us suppose a sequence in which a goal leads to some planned action which through some causal instrumentality determines an outcome. Consider first the case in which this sequence has progressed up to a certain point in reality, and then there are alternative imagined construals, one hedonically positive, the other negative, of the uncertain future course of the sequence. If goal, action, and causal instrumentality are fixed, but only the outcome is uncertain, there is a minimal range of content for the alternative construals to deal with. The associated affective experience can be characterized as SUSPENSE. If goal and action are fixed, but there are alternative causal instrumentalities each potentially controlling the outcome, the alternative construals are inferentially richer and the affect in general will be more elaborate. Perhaps there are alternative authorities who may become responsible for the outcome, as for example two judges who might be assigned to your legal case, one probably sympathetic, the other probaly unsympathetic. The affect here is one of the pair of HOPE/FEAR, depending on whether the favorable or unfavorable construal is emphasized.

When only the goal is fixed, but two (or more) well defined and distinct action plans are construable, each with uncertain connection to the important outcome, the associated affective state may be characterized as AGONIZING. Then not even the goal is fixed, but incompatible possible goals can be clearly construed, the affective state is one of CONFLICT.

Consider next the case in which a particular sequence leading to a favorable outcome is construed in imagination, but reality forces an alternative construal in which the outcome is in fact negative. If goal,action, and causal instrumentality are fixed, but the real negative outcome differs from the imagined positive outcome, the state is one of DISAPPOINTMENT. If goal and action are fixed, but the real causal instrumentality producing a positive outcome differs from the imagined causal instrumentality producing a positive outcome, the affective state is one of FRUSTRATION or ANGER. In relation to previously outlined cases, it can be seen that FRUSTRATION represents dashed HOPE, and DISAPPOINTMENT is negatively resolved SUSPENSE.

In a slightly different type of subcase, the negative reality has already occured, but the individual imagines what might have been, by reconstruing the sequence starting at a particular point of departure. The idea "I shouldn't have done what I did; if only I had acted differently, things would have been different", corresponds to a state of EMBARASSMENT or MORTIFICATION. If the recrimination goes all the way back to believing that one has pursued the wrong goal, then the affective state is one of GUILT.

Another set of subcases arises when there is an imagined negative outcome, but the actual outcome is positive. Without going through all the details, suffice to say that depending on the sequential point at which alternativity occurs, the respective affective states of LUCKINESS, GRATITUDE, PRIDE, AND RECTITUDE can be generated.

Finally, there is the case of incompatible construals which arise because one reality is suddenly replaced by another reality. This need have nothing to do with an intentional action sequence, because it can be outside of the individual's control. If the old reality was positive, and the new reality is no longer positive, the affect is SORROW. If the old reality was not negative, but the new reality is negative, the affect is DISTRESS. If the old reality was not positive, but the new reality is replaced by one which is not negative, a state of RELIEF is produced.

I have been perhaps somewhat scruffy in my presentation of this system of 16 affects (albeit I had earlier implied I would try to be neat). It was not my intention here, however, to be complete and well-disciplined, but only to lay out a particular direction of theory and research involving the role of construals in understanding, memory, and affective experience. The conception of a construal function as a system of subjective constraints which may or may not match objective stimulus constraints is, it seems to me, a very important conception. There is no reason why the idea of systems of constraint should be abandoned to cooption by the right wing within cognitive science, which presumes to investigate Mind without reference to minds. Instead, we need in cognitive science a fusion of left and right wings, of subjective and objective, of content and of formalism.

REFERENCE NOTES

1. Abelson, R.P. & Levi, A. Decision-making and decision theory. Chapter in preparation for G. Lindzey & E. Aronson (Eds.), Handbook of social psychology. Reading, Mass.: Addison-Wesley, 1983.

2. Schank, R.C. Dynamic memory: A theory of learning in computers and people. Manuscript in preparation. Yale University, 1981.

3. Cullingford, R. Script application: Computer understanding of newspaper stories. Research Report 116, Dept. of Computer Science, Yale University, 1978.

4. D'Andrade, R. Cultural meaning systems. Paper presented at the SSRC conference on Culture and Cognition, May 1981.

5. Wilensky, R., Collins, A.,. & Orotny, A. A cognitive analysis of affect. Draft of paper. U. Cal. Berkeley, Nov., 1980.

REFERENCES

Abelson, R.P. Social psychology's rational man. In S.O. Benn & G.W. Mortimore (Eds.), Rationality and the social science. London: Routledge & Kegan Paul, 1976.

Abelson, R.P. The psychological status of the script concept. American Psychologist, 1981, 36, 715-729.

Anderson, R., & Pichert, J.W. Recall of previously unrecallable information following a shift in perspective. Journal of Verbal Learning and Verbal Behavior, 1978, 17, 1-12.

Black, J.B. & Wilensky, R. An evaluation of story grammars. Cognitive Science, 1979, 3, 213-230.

Bower, G.H., Black, J.B. & Turner, T.J. Scripts in memory for text. Cognitive Psychology, 1979, 11, 177-220.

Bush, R.R. & Mosteller, F. Stochastic models for learning. New York: Wiley, 1955.

Carbonell, J.G., Jr. POLITICS: Automated ideological reasoning. Cognitive Science, 1978, 2, 27-52.

deRivera, J. A structural theory of the emotions. Psychological Issues, Monograph 40. N.Y.: International Universities Press, 1977.

Einhorn, H. & Hogarth, R. Behavioral decision theory: Processes of judgement and choice. Annual Review of Psychology, 1981, 32, 53-88.

Galambos, J. & Rips, L.J. The representation of events in memory. Paper presented to the Midwestern Pscyhological Association, May, 1979.

Graesser, A. Prose comprehension beyond the word. New York: Springer-Verlag, 1981.

Hammond, K.R. & Summers, D.A. Cognitive dependence on linear and non linear cues. Psychological Review, 1965, 72, 215-224.

Hastie, R. Memory for behavioral information that confirms or contradicts a personality impression. In R. Hastie, T.M. Ostrom, E.B. Ebbesen, R.S. Wyer, D.L. Hamilton, D.E. Carlston (Eds.) Person memory: The cognitive basis of social perception. Hillsdale N.J.: Erlbaum, 1977.

Keil, F.C. Constraints on knowledge and cognitive development. Psychological Review, 1981, 88, 197-227.

Kintsch, W., Miller, J., & Polson, P. Problems of methodology in cognitive science. Paper for the Cognitive Science Symposium, Boulder, Col., July 1981

Krauss, R.M., Geller, V., & Olson, C.T. Modalities and cause in the detection of deception. Paper presented at the American Psychological Association convention. Washington, D.C., 1976.

Lehnert, W. Text processing effects and recall memory. Research Report 157, Dept. of Computer Science, Yale University, May 1979.

Luchins, A. Mechanization in problem solving. Psychological Monographs, 1942, 54, Whole No. 248.

Mandler, J.M. & Johnson, N.S. Remembrance of things parsed: Story structure and recall. Cognitive Psychology, 1977, 9, 111-151.

March, J.G. Bounded rationality, imbiguity, and the engineering of choice. Bell Journal of Economics and Management Science, 1978, 9, 587-608.

Nisbett, R. & Ross, L. Human inference: strategies and shortcomings of social judgement. Englewood Cliffs, N.J.: Prentice-Hall, 1980.

Roseman, I. Cognitive aspects of emotion and emotional behavior. Paper delivered at the American Psychological Association convention. New York, 1979.

Ross, L., Lepper, M.R. & Hubbard, M. Perserverance in self perception and social perception: Biased attributional processes in the debriefing paradigm. Journal of Personality and Social Psychology, 1975, 32, 880-892.

Rumelhart, D.E. Understanding and summarizing brief stories. In D. LaBerge & S.J. Samuels (Eds.), Basic processes in reading: Perception and comprehension. Hillsdale, N.J.: Erlbaum, 1977.

Schank, R.C. The structure of episodes in memory. In D.G. Bobrow & A. Collins (Eds.), Representation and understanding: studies in cognitive science. New York: Academic Press, 1975.

Schank, R.C. Language and memory. Cognitive Science, 1980, 4, 243-284.

Schank, R.C. and Abelson, R.P. Scripts, plans, goals and understanding. Hillsdale, N.J.: Erlbaum, 1977.

Spiro, R.J. & Esposito, J. Superficial processing of explicit inferences in text. Technical Report 60, Center for the Study of Rading, Urbana, Ill., 1977.

Tomkins, S. The psychology of being right -- and left. Trans-action, 1965, 3, 23-27

Tukey, J.W. Analyzing data: Sanctification or detective work? American Psychologist, 1969, 24, 83-91.

Tversky, A. & Kahneman, D. The framing of decisions and the psychology of choice. Science, 1980.

Wilensky, R. Why John married Mary: Understanding stories involving recurreing goals. Cognitive Science, 1978, 2, 235-266.

SYMPOSIUM GOALS

A Model for Planning in Everyday Situations*

Robert Wilensky

Computer Science Division
Department of EECS
University of California, Berkeley
Berkeley, California 94720

1. Introduction - Four Tenets for a Theory of Planning

Much previous work on planning and problem solving has been concerned with either very specialized systems or with highly artificial domains (e. g., consider Fikes and Nilsson (1971), Newell and Simon (1972), Sussman (1975), Shortliffe (1976)). More recently, there has been an increase in attention given to planning in commonplace situations. For example, Rieger (1975) has proposed a set of "common sense algorithms" for reasoning about everyday physical situations; Hayes-Roth and Hayes-Roth (1979) are concerned with how a person might schedule a day's activities; and Carbonell (1980) POLITIC's program reasons dogmatically about political decisions. On another front, Sacerdoti (1977) and McDermott (1978), while operating perhaps in the more traditional problem solving context, have proposed some powerful approaches to problem solving in general.

1.1. Everyday Planning is Reasoning about Interactions Between Goals

We have been developing a theory of planning that is concerned with reasoning about everyday situations. A central tenet of this theory is that most of the planning involved in everyday situations is primarily concerned with the interactions between goals. That is, planning for individual goals is assumed to be a fairly simple matter, consisting primarily of the straightforward application of rather large quantities of world knowledge. The complexity of planning is attributed to the fact that most situations involve numerous goals that interact in complicated ways.

Thus while traditional problem solving research has been concerned with finding the solution to a single, difficult problem (e. g., finding the winning chess move), most everyday problem solving consists of synthesizing solutions to fairly simple, interacting problems. For example, a typical everyday situation that involves the sort of planning we are interested in might be to obtain some nails, and also buy a hammer. The plan for each goal is straightforward: One simply goes to the hardware store, buys the desired item, and returns. The problem lies in recognizing that it is a terrible idea to execute these plans independently. Rather, the seemingly simple common sense plan is to combine the two individual plans, resulting in the plan of going to the hardware store, buying both items, and then returning.

Simple as this situation may be, most conventional planners are ill-equipped to handle it. Although some planning programs have mechanisms for removing redundancies from a plan, they generally lack a mechanism for even noticing this sort of interaction if these plans are derived from heretofore unrelated goals. Perhaps more importantly, the interaction between plans may have more complex ramifications. For example, if enough items are to be purchased at the hardware store, then a better plan might be to take one's car, while walking may do otherwise. Thus a good part of planning involves detecting the interactions between goals, figuring out their implications, and then deciding what to do about them.

1.2. Planning Knowledge Should be Equally Available for Understanding

The second tenet of our theory of planning is that it should be equally usable by both a planner and an understander. That is, while a planner uses its planning knowledge to bring about a desired state of affairs, an understander may need to use this same knowledge to comprehend the actions of a person it is watching or of a character about whom it is reading. For example, a planner with the goal of keeping fit might take up jogging; an understander might use the same knowledge to infer that someone who has taken up jogging may have done so because he had the goal of staying in shape. Planning and understanding are rather different processes, and this will of course be reflected in our planning and understanding mechanisms. However, our theory of planning specifies that knowledge should be represented in a fashion so that it is usable by either mechanism.

1.3. Meta-Planning is Used as the Driving Principle

The third salient feature of our theory is that it is based on *meta-planning*. By this I mean that the problems a planner encounters in producing a plan for a given situation may themselves be formulated as goals. These "meta-goals" can then be submitted to the planning mechanism, which treats them just like any other goals. That is, the planner attempts to find a "meta-plan" for this meta-goal; the result of successful application of this plan will be the solution to the original planning problem.

A typical example of a meta-goal is the goal RESOLVE-GOAL-CONFLICT. A planner would presumably have an instance of this goal whenever it detects that some of its "ordinary" goals are in conflict with one another. The meta-plans for this goal are the various goal conflict resolution strategies available to the planner.

Meta-planning is described in more detail in Wilensky (1980). Here, we give only a brief characterization of its main features and advantages.

Meta-goals are organized by *meta-themes*. These are very general principles of planning that describe situations in which meta-goals come into being. We summarize these briefly:

Meta-themes

1) DON'T WASTE RESOURCES
2) ACHIEVE AS MANY GOALS AS POSSIBLE
3) MAXIMIZE THE VALUE OF THE GOALS ACHIEVED
4) AVOID IMPOSSIBLE GOALS

*Research sponsored in part by the National Science Foundation under grant MCS79-06543 and by the Office of Naval Research under contract N00014-80-C-0732.

As an example of how these function, the meta-theme "ACHIEVE AS MANY GOALS AS POSSIBLE" is responsible for detecting goal conflicts. That is, if the planner intends to perform a set of actions that will negatively interact with one another, this meta-theme causes the planner to have the goal of resolving the conflict. If this meta-goal fails, i. e., the planner could not find a way to resolve the conflict, then the meta-theme "MAXIMIZE THE VALUE OF THE GOALS ACHIEVED" springs into action. This meta-theme sets up the goal of arriving at a scenario in which the less valuable goals are abandoned in order to fulfill the most valuable ones. The details of the meta-plans involved in these processes are described in length in the last two sections of this paper.

Meta-planning has a number of advantages over other approaches to planning; these advantages are summarized below:

1.3.1. Meta-planning knowledge can be used for both planning and understanding

As meta-goals and meta-plans are declarative structures in the same sense as are ordinary goals and plans, they may be used to understand situations as well as plan in them. Thus an understander with access to this knowledge would be able to interpret someone's action as an attempt to resolve a goal conflict, for example. In contrast, planning programs that have the equivalent knowledge embedded procedurally would not be able to conveniently use it to explain someone else's actions.

1.3.2. The same planning mechanism can apply to more difficult tasks.

Meta-planning knowledge generally embodies a set of strategies for complicated plan interactions. By formulating this knowledge in terms of goals and plans, the same planning architecture that already exists for simpler planning can be used to implement more complicated planning involving multiple goals, etc.

1.3.3. More general resolution of traditional planning problems

Traditional planners usually treat problems such as goal conflicts by special purpose means by the introduction of critics, for example (Sussman 1975, Sacerdoti 1977). This is equivalent to having the general problem solver consult an expert when it gets in trouble. The meta-planning allows the general problem solver to call a general problem solver (itself) instead. Thus all the power of such a system can be focussed on planning problems, rather than just relying on a few expert tactics. Of course, all the specific knowledge usually embodied in critics would still be available to the general problem solver. But the meta-planning model allows this knowledge to interact with all other knowledge as it now take take part in general reasoning processes.

1.3.4. Representational advantages

The meta-planning model also provides more flexibility when no solution can be found. Since a meta-goal represents the formulation of a problem, the existence of the problem may be dealt other than its being fully resolved. For example, the problem solver may simply decide to accept a flawed plan if the violation is viewed as not being too important, or decide to abandon one of the goals that it can't satisfy. By separating solving the problem from formulating the problem, the problem may be accessed as opposed to treated, an option that most other problem solving models do not allow.

1.4. Projection is Used to Infer Goals and Debug Plans

The fourth significant feature of our planning model is that it is based on *projection*. That is, as the planner formulates a plan for a goal, the execution of this plan is simulated in a hypothetical world model. Problems with proposed plans may be detected by examining these hypothetical worlds.

Projection not only enables the planner to find problems with its own plans, but it also enables it to determine that a situation merits having a new goal. For example, sensing an impending danger requires the planner to project from the current state of affairs into a hypothetical world which it finds less desirable. Having done this projection, the planner can infer that it should have the goal of preventing the undesirable state of affairs from coming into being.

Projecting hypothetical realities also allows a general "goal detection" mechanism to work for meta-goals as well as for "ordinary" goals. When proposed plans for goals are projected, interactions will appear in the hypothetical world. Since such interactions generally indicate that some important planning principle is not being adhered to, the occurrence of this hypothetical negative interaction is usually a signal to the planner to achieve some particular meta-goal.

Working with projected universes entails some liabilities as well, as does the notion of meta-planning and of using highly declarative representations. However, our claim is that the prices associated with these ideas are prices that must be paid anyway. By putting them together in the manner described here, a deal of power is obtained for no additional cost.

In the next section, I discuss the general structure of a planning mechanism based on these assumptions. This is the structure used in PANDORA (Plan ANalysis with Dynamic Organization, Revision, and Application), a planning system now under construction at Berkeley. The sections following show how these mechanisms function together in reasoning about *goal conflict situations*. As we have noted, we intend these ideas to be applicable to understanding as well as planning, and in fact, they are being used in a new implementation of PAM, a plan-based story understanding system. While we do not discuss the structure of PAM here, the analysis of goal conflicts is presented in a form in which its use in understanding as well as planning may be seen.

2. The Design of a Planner Based on Meta-planning

This section described the overall architecture of a planner based on the four tenets just considered. The planner is composed of the following major components:

1) The Goal Detector

This mechanism is responsible for determining that the planner has a goal. The goal detector has access to the planner's likes and dislikes, to the state of the world and any changes that may befall it, and to the planner's own internal planning structures and hypothetical world models. The goal detector can therefore establish a new goal because of some change in the environment, because such a goal is instrumental to another goal, or in order to resolve a problem in a planning structure that arises in a hypothetical world model.

2) The Plan Generator

The plan generator proposes plans for the goals already detected. It may dredge up stereotyped solutions, it may edit previously known plans to fit the current situation, or it may create fairly novel solutions. The plan generator is also responsible for expanding high-level plans into their primitive components to allow execution.

3) The Executor

The executor simply carries out the plan steps as proposed by the plan generator. It is responsible for the detection of errors, although not with their correction.

The importance of the goal detector should be emphasized. Most planning systems do not worry about where their goals come from; high-level goals are generally handed to the planner in the form of a problem to be solved. However, a planning system needs to infer its own goals for a number of reasons: an autonomous planner needs to know when it should go into action; for example, it should be able to recognize that it is hungry or that its power supply is low and what goal it should therefore have. It should be able to take advantage of opportunities that present themselves, even if it doesn't have a particular goal in mind at the time. It should be able to protect itself from dangers from its environment, from other planners, or from consequences of its own plans.

The goal detector operates through the use of a mechanism called the *Noticer*. The Noticer is a general facility in charge of recognizing that something has occurred that is of interest to some part of the system. The Noticer monitors changes in the external environment and in the internal states of the system. When it detects the presence of something that it was previously instructed to monitor, it reports this occurrence to the source that originally told it to look for that thing. The Noticer can be thought of as a collection of IF-ADDED demons whose only action is to report some occurrence to some other mechanism.

Goals are detected by having themes and meta-themes asserted into the Noticer with orders to report to the goal detector. *Theme* is a term used by Schank and Abelson (1975) to mean something that gives rise to a goal; a meta-theme, similarly, is responsible for generating meta-goals. For example, we can assert to the noticer that when it gets hungry (i. e., when the value of some internal state reaches a certain point), the planner should have the goal of being not hungry (i. e., of changing this value), or that if someone is threatening to kill the planner, that the planner should have the goal of protecting its life. On the meta-level, we might assert that if a goal conflict comes into existence, then the planner should have the meta-goal of resolving this conflict.

Note that the presumption of a goal detector coupled with meta-planning creates a system of considerable power. For example, no separate mechanism is required for detecting goal conflicts or for noticing that a set of proposed plans will squander a resource. The need to resolve conflicts or conserve resources is expressed by formulating descriptions of the various situations in which this may occur, and the appropriate meta-goal to have when it does. By asserting these descriptions into the Noticer to detect meta-goals, goal conflicts and other important goal interactions are handled automatically.

The planner component of our model itself consists of three components:

1) The Proposer,
   which suggests plausible plans to try
2) The Projector,
   which tests plans by building hypothetical world models of what it would be like to execute these plans
3) The Revisor,
   which can edit and remove parts of a proposed planning structure

The Proposer begins by suggesting the most specific plan it knows of that is applicable to the goal. If this plan is rejected or fails, the Proposer will propose successively more general and "creative" solutions. Once the Proposer has suggested a plan, the Projector starts computing what will happen to the world as the plan is executed. The difficult problems in conducting a simulation involve reasoning about "possible world" type situations which are not amenable to standard temporal logic (McCarthy and Hayes, 1969). However, we finesse this issue by defining hypothetical states in terms of what the planner thinks of in the course of plan construction. In other words, our solution is to let the system assert the changes that would be made into a hypothetical data base, in the meantime letting the goal detector have access to these states. Thus if the plan being simulated would result in the planner dying, say, this would constitute a hypothetical undesirable state, which might trigger further goals, etc.

As the Projector hypothetically carries out the plan, and other goals and meta-goals are detected by the goal detector, the original plan may have to be modified. This is done by explicit calls to the Revisor, which knows the plan structure and can make edits or deletions upon request. The modified plan structure is simulated again until it is either found satisfactory or the entire plan is given up and a new one suggested by the Proposer.

Actually, the function of the Projector is somewhat more pervasive than has so far been described. The Projector must be capable of projecting current events into future possibilities based both on the intentions of the planner and on its analysis of those events themselves. For example, if the planner sees a boulder rolling down the mountain, it is the job of the Projector to project the future path that the boulder will traverse. If the projected path crosses that of the planner, for example, a preservation goal should be detected. Thus the Projector is a quite powerful and general device that is capable of predicting plausible futures.

3. Reasoning about Goal Conflicts

In the next two sections we give a more detailed analysis of one particular part of our planning model, namely, the resolution of goal conflicts. The problem is important in its own right; however, the presentation that follows is aimed at demonstrating the kind of "strategy architecture" to which the model is conducive. In particular, the section illustrates a number of important meta-goals and the meta-plans for them, and describes how they would be invoked and utilized by the model. The section also emphasizes the utility of meta-planning for the application of planning knowledge to understanding goal conflicts as well as to planning for them.

Since it is desirable to achieve all of one's goals, a planner faced with a goal conflict will probably attempt to resolve that conflict. We express this by saying that the state of having a goal conflict is a situation that causes the meta-theme "ACHIEVE AS MANY GOALS AS POSSIBLE" to become active. In such a situation, this meta-theme creates the meta-goal RESOLVE-GOAL-CONFLICT. This is a meta-goal because resolving the conflict can be viewed as a planning problem that needs to be solved by the creation of a better plan. In this formulation, the resolution of the goal conflict is performed by the execution of a meta-plan, the result of which will be a set of altered plans whose execution will not interfere with one another.

The knowledge needed to replan around a goal conflict is quite diverse, and may depend upon the particular goals in question and on the nature of the conflict. However, the meta-plans with which this knowledge is applied are rather general. To see why, it is necessary to ask how it is possible for goal conflicts to be resolved at all. There appear to be two ways in which goal conflicts can come about that determine how they may be resolved:

1) The conflict detected is based on the plans for one's goals, rather than on the goals themselves. In this case, it may be possible to achieve the goals by other, non-conflicting plans.
2) The conflict depends upon some additional circumstance or condition beyond the stated goals or plans. The conflict might therefore be resolved if this circumstance is changed.

We therefore define two very general meta-plans, RE-PLAN and CHANGE-CIRCUMSTANCE. Of course, to be effective, we need to supply these meta-plans with more information; if we use RE-PLAN blindly, for example, we might end up enumerating all possible plans for each conflicting goal, although many of these plan combinations will present the same problem that caused the original goal conflict.

3.1. RE-PLAN

There are a number of different re-planning strategies applicable to goal conflict situations. They are given here in order of decreasing specificity. This is in accordance with our belief about the order in which such plans would actually be used, i. .e, the most specific one first, then progressively more general ones, until a satisfactory set of plans is found. In this respect, meta-plans are entirely analogous to ordinary plans insofar as the planning process is concerned.

The order of plan application is just a corollary of the First Law of Knowledge Application "Always use the most specific piece of knowledge applicable"

3.1.1. USE-NORMAL-PLAN applied to resolving goal conflicts

The most specific re-planning strategy is likewise analogous to the planning strategy for ordinary goals, namely, find a normal plan. A normal plan in the case of goal conflict is to find a stored plan specifically designed for use in a goal conflict between the kinds of goals found in the current situation. For example, consider the following situation:

(1) Mary was very hungry, but she was trying to lose some weight. She decided to take a diet pill.

In (1), there is a conflict between the goal of losing weight and satisfying hunger, as the normal plan for the latter goal involves eating. The RE-PLAN meta-plan is used, and the USE-NORMAL-PLAN strategy applied. The normal plan found that is applicable to both goals is to take a diet pill.

Just as many objects are functionally defined by the role they play in ordinary plans, so some objects are functionally defined by the role they play in plans aimed at resolving specific goal conflicts. Thus a diet pill is an object functionally defined by its ability to resolve the conflict between hunger and weight loss; a raincoat is defined by the role it plays in preventing wetness when one must go outside. In fact, a great deal of mundane planning knowledge appears to consist of plans for resolving specific types of goal conflicts.

3.1.2. Intelligent use of TRY-ALTERNATE-PLAN to find non-conflicting set

A general planning strategy that is applicable when a plan cannot be made to work is to try another plan for that goal. In the case of resolving goal conflicts, this means that alternative plans for each conflicting goal can be proposed until a set is found that are not in conflict. As noted above, this may be costly, but it will only be tried when no canned conflict resolution plan has been found. Moreover, the plan can provide some intelligent ways of proposing new alternatives that may help keep costs down.

For example, consider the following situation:

(2) John was going outside to pick up the paper when he noticed it was raining. He looked for his raincoat, but he couldn't find it. He decided to get Fido to fetch the paper for him.

Here, John first thought to walk outside, but then found that this would cause a conflict. As his normal plan for resolving this conflict failed, John tried proposing other plans, looking for ones that wouldn't entail his getting wet. Since getting the dog to fetch the paper is such a plan, and since John presumably doesn't care if Fido gets wet, this plan is adopted.

The meta-planning strategy used here is called TRY-ALTERNATIVE-PLAN. The difference between using this meta-plan and blind generate and test strategies is that some control can be exerted here over exactly what is undone and and what is looked for as a replacement. For example, the backtracking here need not be chronological or dependency-directed, but can be *knowledge-directed*. That is, rather than undo the last planning decision, a planning decision related to either goal can be undone, possibly based on an informed guess.

In addition, when fetching a new plan, it may be possible to specify in the fetch some conditions that the fetched plan may have to meet without actually testing that plan for a conflict. For example, in the case of getting the newspaper when it is raining, we can ask for a plan for getting something that doesn't involve going outside. That is, we can look for a plan for one goal that does not contain an action that led to the original

conflict. If our memory mechanism can handle such requests, then we can retrieve only those plans that do not conflict in the same way that the original plan does.

In order for this to work, TRY-ALTERNATIVE-PLAN needs to know what part of a plan contributed to the goal conflict so it can look for a plan without this action. This generally depends upon the kind of conflict. We can formulate this within the meta-planning framework by defining a meta-plan called MAKE-ATTRIBUTION. Here, MAKE-ATTRIBUTION is used as a subplan of the TRY-ALTERNATIVE-PLAN meta-plan, although we shall make other uses of it below. TRY-ALTERNATIVE-PLAN first asks MAKE-ATTRIBUTION to specify a cause of the problem, and then fetches a new plan without the objectional element in it.

TRY-ALTERNATIVE-PLAN can also control how far up the proposed goal-subgoal structure it should go to undo a decision. Thus, if no alternative plan for a goal can be found, the goal itself can be questioned if it is a subgoal of some other plan. For example, consider the following scenario:

(3) John was going to get the newspaper when he noticed it was raining. He decided to listen to the radio instead.

Here the entire subgoal of getting the newspaper was eliminated. Since this was apparently a subgoal of finding out the news, the alternative plan of listening to the radio can be substituted. Once again, MAKE-ATTRIBUTION is used to propose a plan that doesn't involve an unwanted step. The difference between this and the last case is that here a plan lying above the conflicting goal is re-planned.

3.2. CHANGE-CIRCUMSTANCE

In addition to the RE-PLAN meta-plan, the other general goal conflict resolution strategy is to change the circumstance that contributes to the conflict. This is actually more general than RE-PLAN, because it may be applicable to conflicts where the goals themselves exclude one another, whereas RE-PLAN requires the conflict to be plan-based.

CHANGE-CIRCUMSTANCE can resolve a goal conflict by altering a state of the world that is responsible for the goals conflicting with one another. Once this has been achieved, the original set of plans may be used without encountering the original problem.

For example, consider the following situations:

(4) John had a meeting with his boss in the morning, but he was feeling ill and wanted to stay in bed. He decided to call his boss and try to postpone the meeting until he felt better.

(5) John wanted to live in San Francisco, but he also wanted to live near Mary, and she lived in New York. John tried to persuade Mary to move to San Francisco with him.

In (4), John's conflict is caused by his plan to attend the meeting and his plan to stay home and rest. These plans conflict because of the time constraints on John's meeting, which force the plans to overlap; the plans require John to be in two places at once, so they cannot be executed simultaneously. If the time constraint on attending the meeting were relaxed, however, then the conflict would cease to exist. Thus rather than alter his plans, John can seek to change the circumstances that cause his plans to conflict by attempting to remove the time constraint that are a cause of the difficulty.

In (5), the conflict is between living in San Francisco and being near Mary, who is in New York. The basis for this exclusion involves the location of San Francisco and of Mary. However, if one of these locations were changed so that the distance between them were reduced, them the state would no longer exclude one another. Thus John can attempt to change Mary's location, while still maintaining his original goals.

To decide what circumstance to change, a planner once again needs to analyze the cause of the conflict. Thus CHANGE-CIRCUMSTANCE first requires the use of MAKE-ATTRIBUTION to propose a candidate for alteration. As was the case for RE-PLAN, MAKE-ATTRIBUTION requires access to detailed knowledge about the nature of negative goal interactions in order to find a particular circumstance with which to meddle. An analysis of such interactions appears in Wilensky (1978).

4. Goal Abandonment

When attempts to resolve a goal conflict are unsuccessful, a planner must make a decision about what should be salvaged. In terms of meta-planning, we can describe these "goal abandonment" situations as follows. The inability to achieve a RESOLVE-GOAL-CONFLICT meta-goal results in the planner having this failed meta-goal. Having a failed RESOLVE-GOAL-CONFLICT meta-goal is a condition that triggers the meta-theme MAXIMIZE THE VALUE OF THE GOALS ACHIEVED. This triggering condition causes this meta-theme to invoke a new meta-goal, called CHOOSE-MOST-VALUABLE-SCENARIO. This goal is satisfied when the relative worth of various achievable subsets of the conflicting goals is assessed, and the subset offering the greatest potential yield determined.

To achieve this meta-goal, we postulate a SIMULATE-AND-SELECT meta-plan. This plan proposes various combinations of goals to try, and computes the worth of each combination. The most valuable set of goals is returned as the scenario most worth pursuing.

4.1. The SIMULATE-AND-SELECT meta-plan

The SIMULATE-AND-SELECT meta-plan has a rich structure. To begin with, it makes a number of presumptions about evaluating the cost and worth of goals and of comparing them to one another. We presume that values can be attributed to individual states in isolation *ceteris paribus*, and that the value of a set can be computed from its parts. This does not presume that the computation is simple; indeed, it may involve the consultation of large amounts of world knowledge. However, we do assume that all values can be made commeasurable. The general issues involving attributing values to goals are discussed (although by no means resolved) in Wilensky (1980).

The SIMULATE-AND-SELECT meta-plan has in effect two distinct options. The first is quite straightforward. It simply involves constructing maximal achievable (i. e., non-conflicting) subsets from among the conflicting goals, and evaluating the net worth of each one. Since we are generally dealing with two goals in a conflict, this means just evaluating the worth of one goal and comparing it to the value of the other. Thus if having the newspaper is deemed more valuable than getting wet, then the planner walks outside to get the newspaper and allows himself to get soaked. Alternatively, a reader trying to understand someone else's behavior would use knowledge about this meta-plan to make inferences about their value system. If we observe John risking getting wet into to get his morning paper, then we conclude that having his paper is worth more to him than getting wet.

However, there is another set of alternatives that need to be considered. Consider once again the example of fetching the newspaper in from the rain, in which the original goals are to get the newspaper and to remain dry. Rather than abandon either goal completely, a reasonable alternative is to try to reduce the degree to which one gets wet as much as possible. A plan for remaining as dry as possible while moving through the rain is to run as fast as one can. This plan satisfies one goal entirely, and another to a degree. The total value of this scenario is likely to be greater than the value of

staying dry but not getting the paper, and since the other abandonment possibility (getting the paper but getting soaked) is clearly worse than this (i. e., getting the paper but getting less soaked), the scenario involving partial fulfillment is likely to be adopted.

Partial goal fulfillment is a general principle that is applicable to all goals that involve scalar values. It allows the SIMULATE-AND-SELECT meta-plan to propose options in which the partial fulfillment of one goal enables the (possibly partial) fulfillment of the other. This process is illustrated by the "newspaper in the rain" example: MAKE-ATTRIBUTION determines that the problem with the "stay dry" goal above is that it requires not going outside. Thus a partial version of this goal is sought that doesn't involve this condition. In the case of not getting wet above, the "stay as dry as possible" alternative is selected because this doesn't require not going outside. This scenario is therefore hypothesized and evaluated along with the strict abandonment options, and the one with the highest value chosen.

5. Summary and Projections

We have proposed a theory of planning based on four tenets: (1) Commonsense planning is essentially the consideration of interactions of otherwise simple plans, (2) knowledge about planning should be usable both by a planner and an understander, (3) planning problems should be formulated as meta-goals, and solved by the same planning mechanism responsible for the fulfillment of ordinary goals, and (4) to accomplish much of its mandate, the planner makes projections of the future based on its current knowledge of the world and its own tentative plans.

These tenets form the basis for a model of planning whose most salient features are a goal detector and a projector. The goal detector is used to infer goals, including meta-goals, based on the situations in which the planner finds itself; the projector is used to guess what the future will bring based on the planner's current beliefs and plans. As the goal detector has access to the hypothetical situations simulated by the projector, it can detect problems with currently intended plans by noticing their consequences in hypothetical realities. These problems are dealt with by setting up meta-goals to try to assure a more desirable future state of affairs.

We examined this model of planning in the particular domain of goal conflict resolution. Here we found use for the meta-plans RE-PLAN (consisting of USE-NORMAL-PLAN and USE-ALTERNATE-PLAN) and CHANGE-CIRCUMSTANCE for the meta-goal RESOLVE-GOAL-CONFLICT. Both meta-plans make use of the powerful sub-plan MAKE-ATTRIBUTION. For the related goal of CHOOSE-MOST-VALUABLE-SCENARIO, the SIMULATE-AND-SELECT meta-plan is used to create alternatives involving goal abandonment and partial goal fulfillment. MAKE-ATTRIBUTION was found to be useful here as well.

We are currently attempting to test these ideas in two programs. PAM, a story understanding system, uses knowledge about goal interactions to understand stories involving multiple goals. That is, PAM can detect situations like goal conflict and goal competition, and, realizing that these threaten certain meta-goals, PAM will interpret a character's subsequent behavior as a meta-plan to address the negative consequences of these interactions. Thus PAM makes use of the knowledge structures described above, but of course, it does not test the model of planning per se.

Both the model of planning knowledge and of planning is being used in the development of PANDORA (Plan ANalyzer with Dynamic Organization, Revision and Application). PANDORA is given a description of a situation and determines if it has any goals it should act upon. It then creates plans for these goals, using projection to test them. New goals, including meta-goals, may be inferred in the process, possibly causing PANDORA to revise its previous plans

The following is an example of the kind of planning situation that PANDORA can handle. PANDORA is presented with a task that requires it to get some nails and to get a hammer. PANDORA proposes the normal plans for these goals, which require it to go to the store, buy the desired item, and return. As the plans involve some common preconditions, the meta-theme "DON'T WASTE RESOURCES" causes PANDORA to have the meta-goal COMBINE-PLANS. A meta-plan associated with this goal synthesizes a new plan that involves going to the store, buying both objects, and returning.

PANDORA can also detect and resolve a number of goal conflict-base situations. In addition, PANDORA is being used to model the planning processes of a human who needs to cook dinner during a power failure, in which most of the normal plans for one's goals will not be effective.

References

1 Fikes, R. and Nilsson, N. J. (1971). STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence 2*, 189-208.

2 Hayes-Roth, B. and Hayes-Roth, F. (1978). Cognitive Processes in Planning. RAND Report R-2366-ONR.

3 McDermott, Drew (1978). Planning and Acting. In *Cognitive Science* vol. 2, no. 2.

4 McCarthy, J. and Hayes, P. J. (1969) Some philosophical problems from the standpoint of artificial intelligence. In Meltzer and D. Michie (eds.) *Machine intelligence*, vol. 4. New York: American Elsevier, pp. 463-502.

5 Newell, A., and Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, N. J.: Prentice Hall.

6 Rieger, C. (1975). The Commonsense Algorithm as a Basis for Computer Models of Human Memory, Inference, Belief, and Contextual Language Comprehension. In *Theoretical Issues in Natural Language Processing*, R. Schank and B. L. Nash-Webber, (eds.). Cambridge, Mass.

7 Sacerdoti, E. D. (1977). *A Structure for Plans and Behavior*. Elsevier North-Holland, Amsterdam.

8 Schank, R. C. and Abelson, R. P. (1977). *Scripts, Plans, Goals, and Understanding*. Lawrence Erlbaum Press, Hillsdale, N.J.

9 Shortliffe, E. H. (1976). *MYCIN: Computer-based Medical Consultations*. American Elsevier.

10 Stefik, Mark J. (1980). Planning and Meta-Planning MOLGEN: Part 2. Stanford Heuristic Programming Project HPP-80-13 (working paper), Computer Science Department, Stanford University.

11 Sussman, G. J. (1975). *A Computer Model of Skill Acquisition*. American Elsevier, New York.

12 Wilensky, R. (1978). Understanding goal-based stories. Yale University Research Report No. 140.

13 Wilensky, R. (1980). Meta-planning: Representing and Using Knowledge about Planning in Problem Solving and Natural Language Understanding. Berkeley Electronic Research Laboratory Memorandum No. UCB/ERL/M80/33.

A Commitment-Based Framework for Describing Informal Cooperative Work

by

Richard E. Fikes
Cognitive and Instructional Sciences Group
Xerox Palo Alto Research Center

June 1981

Abstract

In this paper we present a framework for describing cooperative work in informal domains such as an office. We argue that standard models of such work are inadequate for describing the adaptibility and variability that is observed in offices, and are fundamentally misleading as metaphors for understanding the skills and knowledge needed by computers or people to do the work. The basic claim in our alternative framework is that an agent's work is defined in terms of making and fulfilling commiunents to other agents. The tasks described in those commitments are merely agreed upon means for fulfilling the commitments, and the agents involved in the agreement decide in any given situation how and whether a given commitment has been fulfilled. We also claim that in informal domains, descriptions of tasks, functions, and procedures are necessarily incomplete and imprecise. They result from negotiations among the agents and serve as agreed upon specifications of what is to be done. The descriptions evolve during their use in continuing negotiations as situations and questions arise in which their meaning is unclear. These claims imply that for a given situation an agent using such descriptions must be capable of interpreting imprecise descriptions, determining effective methods for performing tasks, and negotiating with other agents to determine task requirements.

Introduction

In this paper we present a framework for describing cooperative work in domains where there is no agreed upon formal model of the tasks and functions to be done, nor of the procedures for doing them (i.e., in informal task domains). Most of the work people do is in such informal domains where it is not feasible to create precise statements of the tasks to be done, the situations in which they are to be done, the resources available, nor the capabilities of the processors doing them. For example, people are regularly confronted with task descriptions such as "write a progress report on the project", "describe the items to be purchased", "yield right of way", "keep your eye on the ball", or "slice chicken breasts very thinly into julienne strips"

The project that produced this framework has been focused on the problems of automating office work involving the use of prespecified procedures, and our experiences with those problems motivate and provide examples for the discussions in this paper. However, the results reported here are intended to be generally applicable in any task domain where tasks and procedures are informally specified, and agents enlist each other's help to achieve their individual goals.

We began our office automation efforts by attempting to develop a model of the work being automated that would provide a basis for our system design efforts. We were particularly interested in describing the skills and information needed to do the work, and in accounting for the adaptibility and variability of methods used by people in performing their tasks.

Our initial thesis was that office procedures are like computer programs and they are "executed" by a collection of office workers in a manner analogous to a collection of computers executing a program. It seemed a simple matter to automate office procedures by storing them in a computerized data base as if they were programs. Then, at each step in the execution of the procedures in an office the computer could do the step itself, or tell the person doing the step what operation to do and then monitor the results.

As we proceeded, fundamental problems arose that led us to question our thesis [Fikes and Henderson]. One such problem was that our model did not account for the variability in the way tasks are accomplished. For example, an office worker has more options in following a procedure than our model described. He can choose to ignore some of the requirements of a task (e.g., leave some fields of form blank), renegotiate the requirements of a task (e.g., request extension of a deadline), or use some method other than the standard procedure for performing a task.

A second problem was that our model did not account for the difficulties related to working with informally specified tasks, functions, and procedures. For example, the informality of office work makes infeasible the specification of precise algorithms for performing tasks. Situations occur in which the available procedures do not indicate what to do (e.g., a vendor claims that goods were delivered, but no record of their arrival can be found), what is indicated cannot be done (e.g., a deadline has already past), or what is indicated is not the preferred method for performing the task (e.g., because an unexpected resource is available). Hence, the work involved in using office procedures is qualitatively different from the work involved in executing a formal algorithm.

We concluded, then, that modeling procedural office work as simple program execution is an inadequate basis for automating it and is misleading as a metaphor for understanding the skills and knowledge needed by computers or people to do the work. That conclusion led us to search for alternative ways of modeling cooperative work that would account for the way office tasks are actually performed, and would inform us regarding the required skills and knowledge. This paper presents the initial results of that search, an alternative based on the agreements made by agents performing the work and agents for whom the work is done (see [Flores and Ludlow] for another analysis of office work based on such agreements).

Analyzing Cooperative Informal Work

The Social Nature of Tasks and Functions

Tasks

We begin by analyzing a simple work situation in which an agent has a task that he wants done. For the purposes of this discussion we will consider a task to be defined as a set of goals to be achieved while maintaining a set of constraints. The basic tenet of our model is that tasks are essentially social in nature in that they are done by one agent (a contractor) for some other agent (a client). The situation where an agent does a task for himself is the special case in which the client and contractor are the same agent.

A client (i.e., an agent who wants a task done) can choose to enlist some other agent to be the contractor for a task. The client accomplishes that enlistment by establishing a "task contract" with the contractor (see the work on contract nets [Davis and Smith] [Smith] for a detailed model of how such contracts are established). We model a contract as consisting of a collection of commitments, each made by one of the contracting agents to another of the contracting agents. A task contract is an agreement between two agents containing a commitment by one of the agents (the contractor) to do a task for a second agent (the client). The contract may also contain other commitments, such as a commitment by the client to remunerate (i.e., achieve some goal for) the contractor in return for accomplishing the task.

In order for a task contract to be established, the client and contractor must agree on the task that is to be performed. Their negotiations will produce an agreed upon *description* of the task, and the commitment will be a statement of intent to do the *described task*. Therefore, we consider cooperative tasks as being defined by a social process, and as representing a negotiated agreement between the client and contractor.

Note that a task contract establishes a goal for the contractor of fulfilling his commitment to the client, and performing the described task is only a means for achieving that goal. That observation is one of the bases for our explanations of the behavior variations observed in human cooperative work situations. That is, the agreed upon task description provides the contractor with a set of sufficient conditions for fulfilling his commitment, and therefore represents one method of achieving his goal. However, *any* actions by the contractor that result in the client agreeing that the commitment has been fulfilled will achieve the contractor's goal. For example, the contractor may choose to achieve some variation of the task's goal, ignore some of the constraints, convince the client that the task shouldn't be done, etc. He is free to use whatever method he thinks will succeed and is most desirable for him in the context of his other goals and constraints.

In order for a contractor to make use of the flexibility available to him in fulfilling his commitment, he must know who the client is, the client must be accessible to him for negotiation, and the contractor must be capable of planning and performing alternative courses of action to those described in the task contract.

Functions

Often in human work situations, a person will agree to perform a given type of task whenever a given set of conditions occur; that is, he will agree to perform a "function". For example, a buyer in a corporate procurement department may agree to issue a purchase order whenever a properly executed purchase request is received. Also, procedures are typically designed as methods for performing functions rather than individual tasks and are used whenever the function's task is to be done (e.g., the procedure for issuing purchase orders). Hence, in order to describe those situations, we will generalize our discussion to include functions as well as tasks.

As with tasks, one method available to a client for performing a function is to establish an agreement (a "function contract") with a contractor in which the contractor commits to do the function for the client. The contract will contain an agreed upon description of the function to be performed by the contractor. For our discussion, we will consider a function description to consist of a parameterized task description and parameterized set of preconditions such that any given instance of the preconditions defines an instance of the task. Whenever an instance of the preconditions becomes true, the contractor agrees to perform the corresponding instantiated task.

As with tasks, the contractor's goal is to fulfill his commitment and the agreed upon function description provides him with a set of sufficient conditions for achieving that goal. Each time the function's preconditions become true, the contractor can choose to do whatever actions he thinks will satisfy the client.

Note that in the transition from task to function a new subtask has been introduced; namely, recognition of the occurrence of the preconditions. Hence, an agent who has committed to perform a function must establish monitors that recognize situations in which an instance of the function's task is to be performed.

Tasks and Functions in Informal Domains

An agent performing a function depends on the function description to specify each situation in which he is to do something and in each of those situations the task that he is to perform. In informal domains, use of those descriptions becomes problematic because of their inprecision and incompleteness (What is a "properly executed purchase request"?) [Suchman]. Hence, the contractor is confronted with the new subtasks in each situation of interpreting the function description to determine whether a task is to be done, what the task would be, and then after doing something whether the task has been accomplished.

We claim in our model that the sole criteria for an acceptable interpretation of these descriptions is agreement by the contractor and client. That is, the function and task descriptions are part of the contract between the contractor and client, and those agents are the final authority as to what those descriptions mean and whether they have been satisfied. For example, the meaning of "describe the items to be purchased" on a purchase requisition form is worked out in each case by the requisitioner and the procurement department buyer, for whom the description is being created.

The interpretation of task and function descriptions in any given situation is therefore a subject for negotiation between client and contractor. That is, if a commitment description is not sufficiently precise or complete for the contractor to determine what he should do in a given situation, then additional negotiations with the client are necessary. Hence, in informal domains, the negotiation processes that produce commitment descriptions continue during the fulfillment of those commitments and become an integral part of the work required to fulfill them.

Agents performing functions in informal domains must be capable of determining appropriate interpretations of imprecise descriptions and of recognizing when the description is sufficiently inadequate to warrant renegotiation with the client. When agents are skilled in those capabilities, the difficult and time consuming process of creating comprehensive function and procedure descriptions can be avoided. Descriptions can be allowed to build up incrementally by generalizing the experiences gained in particular situations.

Functions as Operators for Planning

Functions play the same role as operators in standard Artificial Intelligence planning and problem solving frameworks (for example, [Fikes and Nilsson]) in that they can be used by agents to achieve goals. We said earlier that an agent who wants a task done can enlist a second agent to do the task by establishing a task agreement with the second agent. Functions provide an alternative method of enlisting a second agent to do a task. That is, if the second agent is a contractor who has made a committment (it doesn't matter to whom) to provide a function, and the task that the first agent ("the consumer") wants done is an instance of that function's task, then the consumer can cause the contractor to do the task by persuading him that the appropriate instance of the function's preconditions are true. If the contractor refuses to do the task, then the consumer can appeal to the function's client, attempting to convince him that the preconditions were satisfied and that the contractor did not fulfill his commitment to accomplish the task.

For example, if an employee of a small company wants to obtain some equipment for use in his work, then he can achieve that goal by obtaining the appropriate authorizations and submitting the appropriate forms to the company's procurement department. The procurement department has made a commitment to the company president to be the contractor for the function of purchasing equipment, and receipt of the appropriate forms and authorizations is the precondition for that function. The employee becomes a consumer of that function by convincing the contractor that an instance of its preconditions have become true. If the procurement department refuses to provide the advance, the employee can complain to the company president that they are not performing their function.

In deciding to use a function, the consumer has replaced his original task with the new task of persuading the contractor to do the original task. Notice that the method for accomplishing the new task is to convince the contractor that an instance of the preconditions have been satisfied, rather than simply to make an instance of the preconditions true. The consumer is free to negotiate with the contractor as to what he will accept as satisfactory evidence that the preconditions are true. For example, the employee requesting an equipment purchase might convince the procurement department that a phone call from the employee's manager is sufficient in that case to authorize the purchase. If the preconditions are informally described, then there is the additional issue to be resolved in those negotiations of determining an agreed upon interpretation of the descriptions in the situation. For example, the employee might ask the procurement department to accept a memo requesting the purchase rather than the standard form. This is another case where negotiations during the performance of a task are vital to its completion and where variability is introduced by the one-time agreements that result from those negotiations.

Subcontracting to Perform Tasks and Functions

Consider again the basic situation in which a client wants a task done and has obtained a commitment from a contractor to do the task. We could then describe the contractor's situation as one in which he wants a task done, and that he has the option of persuading yet a third agent (a "subcontractor") to do some or all of the task for him. The subcontractor then is in the same situation and has an option to enlist a fourth agent, etc. The same description holds for functions as well as tasks.

We are interested here in examining the role that the contractor's client plays in the work of a subcontractor. For that purpose it is sufficient to consider the three agent case where a contractor and client have an agreement in which the contractor commits to perform a function, and the contractor instead of performing the function himself establishes an agreement with a subcontractor in which the subcontractor commits to perform the function. In that case, the contractor's client then becomes the consumer for the subcontractor's function.

We can augment our purchasing example by considering the function contract between the company president and the employee. In that contract, the president commits to purchase equipment for the employee whenever he submits an authorized request. Instead of doing the purchasing himself, the president contracts with the procurement department to do it. Hence, a subcontracting relationship exists in which the employee is the consumer, the company president is the contractor, and the procurement department is the subcontractor. Figure 1 indicates the structure of the two contracts that establish those relationships.

Main Function Contract:

Client: The employee
Contractor: The company president
Function Description: Purchase equipment for the employee whenever he submits an authorized request.

Function SubContract:

Client: The company president
Contractor: The procurement department
Consumer: The employee
Function Description: Purchase equipment for the employee whenever he submits an authorized request.

Figure 1: Example subcontracting situation in an office

The contractor wants the function done in order to fulfill his commitment to the consumer. The commitment of the subcontractor to perform the function can therefore be considered as being to fulfill the contractor's commitment to the consumer. Satisfaction of the consumer is a sufficient condition and important method for the subcontractor to fulfill his commitment. The subcontractor can therefore do whatever he thinks will convince the consumer that the contractor's commitment to him has been fulfilled.

The consumer therefore plays an important role in the subcontactor's work and is an additional agent with whom the subcontractor can negotiate to determine what is required of him in a given situation. As before, if the work is being done in an informal domain, then determining agreed upon interpretations of the descriptions in particular situations is an additional issue for negotiation. If the subcontractor and the consumer agree on what is to be done, then the contractor need not enter into the negotiations or even know what was agreed on because his commitment to the consumer is being fulfilled and the commitment to him by the subcontractor is being fulfilled.

If, in a given situation, the subcontractor and consumer cannot agree on the task to be done, then they both can appeal to the contractor for help. The subcontractor can argue that his commitment to the contractor does not include what the consumer is asking for, and the consumer can argue that the contractor's commitment to him is not being fulfilled. Hence, the contractor needs to enter into the negotiations only when the subcontractor and the consumer cannot agree.

For example, when the employee requests the equipment purchase, the procurement department buyer may attempt to satisfy the employee by convincing him that he should use previously purchased equipment or that he should rent equipment. He may persuade the employee to help find an appropriate vendor, and in return agree to obtain the authorizations for the purchase that are normally part of the employee's responsibility. Such localized one-time agreements between agents occur regularly in office settings, and are an important aspect of the variability and adaptibility that characterize office work. Standard computer program description techniques (e.g., flow charts) are hopelessly inadequate for describing such activity.

So, we see that the consumer is a source of information for the subcontractor about what is to be done and an authority on when the task has been completed. Also, the consumer acts as a monitor for the contractor as to whether the subcontractor has done his job, since it is the consumer who cares whether or not the task is accomplished. The interdependencies among the consumer, contractor, and subcontractor discussed in this section are summarized in Figure 2.

For the consumer:

The subcontractor:
Performs the desired task.
The contractor:
Settles disputes with the subcontractor.

For the contractor:

The subcontractor:
Fulfills the commitment to the consumer.
The consumer:
Provides remuneration for doing the task, and monitors the subcontractor's work.

For the subcontractor:

The consumer.
Helps interpret the task description, and Indicates when the task is completed.
The contractor:
Provides remuneration for doing the task, and helps settle disputes with the consumer.

Figure 2: Summary of the Consumer, Contractor, Subcontractor Relationships

The Social Nature of Procedures

Now consider a situation in which an agent has a function he wants done and a procedure describing how to do it. We will call the agent who has the function "the procedure's manager" and the function "the procedure's function". A procedure describes a method for doing a function in terms of a collection of steps to be done in a specified order, and thereby provides a means for the procedure's manager to organize a collection of agents to perform the procedure's function. That is, the procedure's manager has the option for each step of the procedure of obtaining a commitment from some other agent to do the step (i.e., of "installing the step"). If he obtains such a commitment for each step of the procedure (i.e., if he "installs the procedure"), then the agents who agreed to do the steps (i.e., the "step contractors") will do the function for him. For example, if a procurement department manager is assigned the function of purchasing equipment for employees, then he can either find or create a procedure for performing that function and install the procedure by obtaining commitments from the people in his department to be step contractors for each of the procedure's steps.

In formal domains, operation descriptions can be provided for each step in a procedure that are guaranteed to satisfy the designer's intention for the step (e.g., add x to y). Then the commitment of a step contractor is to perform the step's operation in a manner that satisfies the formal description. The contractor need not have any model of the results expected from his step or of the role they play in performing the procedure's task. His total sphere of concern is to perform the operation as specified. That is the style of procedure execution done, for example, by a typical programming language interpreter.

In informal domains, there are no guarantees that a procedure will successfully accomplish its task. Those guarantees are lost because the procedure, its task, and the situations in which it will be used are imprecisely described. Hence, procedures in informal domains are only prototypes of methods for performing tasks. They suggest a way of decomposing a task into steps, and perhaps indicate how the task is typically performed, but they do not alleviate the need for problem solving in each specific situation to determine how to perform a task. The user of an informal procedure is confronted with the subproblems of determining the meaning of the procedure in the specific situation and whether it will be applicable or effective.

A basic problem in informal domains with installing procedures to perform functions is that one must commit at the time of installation to the decomposition specified by the procedure. If indeed as we argued above, that decomposition is only suggestive and needs to be reexamined each time the procedure is used, then the strategy of installing a procedure is an ineffective means of transfering the work from the procedure's manager to the step contractors. The challenge then is to describe and install procedures in a manner that maximises their adaptibility and flexibility.

Procedure Steps as Functions

An important way of meeting the challenge of compensating for the inadequacies of informally specified procedures is to add to the description of each step a description of the *function* to be accomplished by that step (i.e., the goals to be achieved and constraints to be maintained each time the step is performed). For example, add to a step described as "Submit to procurement an authorized purchase request" the function description "Whenever an employee wants equipment purchased, achieve the goal: Procurement knows the employee wants equipment purchased and has the information and authorizations necessary to make the purchase".

A function description specifies the requirements of a step without reference to *how* those requirements are to be performed and therefore provides the option of using whatever method is appropriate in a particular situation to accomplish the function's task. The agent performing a step can use the function description to evaluate whether the action described for the step is an appropriate method in a given situation, to plan alternative methods for performing the step, and to evaluate whether his actions accomplished the step.

Adding function descriptions to steps results in procedures applicable to a wider range of situations because it allows the agents performing the steps to take into consideration properties of the situation such as resource limitations and interactions with other tasks that may not have been known at the time the procedure was designed. The work involved in using those function descriptions is significantly different from the work of performing steps described as operations. In particular, it involves subtasks of *planning* to determine a method to use, and *monitoring* to determine whether the method accomplished the function. However, an agent capable of effectively performing those subtasks can better determine the appropriateness of his results and can successfully perform his step in unexpected situations [Fikes].

Subcontracting Within Procedures

Our description of procedure installation thus far would predict that each time a procedure step is activated and the step contractor does something other than the task described in his agreement with the procedure's manager, that the contractor must obtain an acknowledgement from the manager that what he did satisfies his commitment. In actual practice in offices, there is a broad variability of behavior in the performance of procedure steps, and only rarely is that behavior accompanied by interaction with the procedure's manager (typically the step contractor's supervisor). Instead, there are frequent negotiations among the agents doing the steps of the procedure. Those agents are not generally working for each other and have made no apparent commitments to each other. How do we explain their negotiations and the role those interactions play in their work? In this section we model those interactions by extending our description of procedure installation to include the subcontracting relationships that are established among the step contractors.

We can apply our analysis of subcontracting to the performance of procedure steps by identifying the commitments made during a procedure installation and considering the "functional role" played by procedure steps. A step's functional role is the rationale used by the procedure designer for including the step in the procedure (e.g., achieve a goal of the procedure's task, satisfy a precondition of some other step in the procedure). That rationale is therefore the defining basis for the function to be performed at that step [VanLehn and Brown].

The function to be performed at each step of a procedure has a set of preconditions as part of its description. The designer of a procedure must assure that when a given step is to be performed, its preconditions are satisfied. That design goal is satisfied by including other steps earlier in the procedure that will cause those preconditions to be satisfied. The functional role of those earlier steps, therefore, is to satisfy the preconditions of the later step.

We can characterize a function's preconditions as consisting of "activation conditions", the occurrence of which signals the contractor that an instance of the function's task is to be done, and "enabling conditions", the satisfaction of which provides the context needed by the contractor to perform the task. For example, the function performed by a buyer in a procurement department is activated when he receives a purchase request and is enabled when he receives authorization to make the purchase. We distinguish, therefore, between steps whose functional role is to *activate* other steps and those whose role is to *enable* other steps.

We make use of that distinction in describing the contract that installs a procedure step. That contract contains a commitment by the step contractor to perform the step's function whenever the step's activation conditions occur and a commitment by the procedure's manager to satisfy the step's enabling conditions whenever the activation conditions occur. For example, an accounting department clerk (the step contractor) may make a commitment to his manager (the procedure's manager) to respond to vendors' invoices whenever one is received. The manager would, in turn, agree to provide the clerk with the purchase order, packing slips, and other supporting documents needed to respond appropriately to the vendor.

The procedure's manager satisfies his commitment to satisfy a step's enabling conditions by installing those procedure steps whose functional role is to enable that step. Hence, an agent who is performing a step whose functional role is to enable some other step is in effect a subcontractor whose consumer is the agent performing the step he is enabling. In the accounting department example above, the agents who supply the clerk with the supporting documents are subcontractors whose consumer is the clerk.

Our earlier comments about the role that a consumer plays in the work of a subcontractor therefore apply here. The agent doing the step being enabled and the agent satisfying the enabling condition negotiate with each other to determine what the enabler's task is in problematic situations, and the procedure's manager is brought into the negotiations only when they cannot agree. Also, the agent being enabled acts as a monitor on the enabler for the procedure's manager.

The analysis of subcontracting applies to any procedure step whose functional role involves providing a result to some agent other than directly to the procedure's manager. In those cases the agent providing the result is fulfilling a commitment made by the procedure's manager to the consumer of that result (or to a client of that consumer). Hence, the consumer and producer can work out together what is to be provided.

We conclude from this discussion that an important way of increasing the adaptibility of a procedure is to include in the description of each step the functional role of the step. If that functional role involves fulfilling a commitment of the procedure's manager to some third agent, then the description should include the identity of that agent, and the step contractor should have access to him for ongoing negotiations.

Summary and Conclusions

In this analysis we have described a framework that identifies the sources of variability and adaptibility observed in human cooperative work situations. Our basic claim is that an agent's work is defined in terms of making and fulfilling commitments to other agents. The tasks described in those commitments are merely agreed upon means for fulfilling the commitments. The agents involved in the agreement are free in any given situation to decide how and whether a given commitment has been fulfilled. Hence, nonstandard methods and outcomes may be considered acceptable even though they do not correspond to the described tasks, functions, and procedures.

We claimed that descriptions of tasks and functions result from negotiations between clients and contractors, and serve as agreed upon specifications of what the contractor is to do. In informal domains, those descriptions are necessarily incomplete and imprecise. Determining their intended meaning in specific situations is an important component of the work. That determination involves continuing negotiations as situations and questions arise in which the meaning of the decriptions is unclear.

Procedures provide a means for organizing a collection of agents to perform a function. In informal domains, procedures represent only prototypes of methods whose meaning and applicability in specific situations is unclear. Their use requires problem solving and negotiation to determine an effective method in a given situation.

Information Needed To Do Cooperative Procedural Work

This framework characterizes the information needed by agents doing cooperative work and the role that the information plays in their work. In general, it indicates that an agent needs descriptions of the task and function contracts to which he has agreed, and the functions available to him.

For each task or function commitment that an agent has made, he needs to know the agreed upon task or function description (because it provides a set of sufficient conditions for fulfilling the commitment), the agent to whom the commitment was made (so that the contractor knows whose satisfaction he is trying to obtain), and the consumer of the results of the task or function in the case where the commitment is a subcontract (because satisfying the consumer is a sufficient condition for fulfilling the commitment).

An agent needs to know the functions available to him so that he can use them as steps in plans he forms to accomplish his tasks. In order to use a function, he needs a description of its task (so that he can determine whether the function can be used to accomplish his task), its preconditions (because they describe a means for initiating performance of the task), the identity of the contractor (so he will know who he must persuade to perform the task), and the identity of the client (so he will know who to appeal to if he feels that the contractor is not adequately performing the function).

Information Needed From a Procedure Description

We have also indicated information that is needed from the description of an informal procedure in order for the procedure to be used adaptively and flexibly. The description should identify the procedure's manager (so that each step contractor knows whose satisfaction he is trying to obtain), and each step of the procedure should be described as a function (so that the step contractor can choose his own method of performing the step). If satisfaction of an enabling condition of a step is subcontracted to another step, then the description of the step being enabled should identify the enabling step and who is performing it (so that the contractor for the enabled step can negotiate with the enabler and monitor his performance). Finally, as noted above about all functions, if a step is a subcontract, then its description should identify the consumer of the subcontract (because satisfying him is a sufficient condition for fulfilling the commitment).

Implications for Office Automation

This framework is serving as a basis for our exploration of how computer-based systems can effectively participate in procedural work in offices. We have reported in earlier papers our preliminary results in this regard ([Fikes] and [Fikes and Henderson]) and will not attempt to describe our current efforts in detail here. Instead, we will conclude this paper with some general remarks on office automation to suggest the uses we are making of the commitment-based framework.

Our discussion indicates that in informal domains, "intelligent" capabilities such as planning, plan monitoring, and negotiation are required to do even simple cooperative work. Current computer-based systems that claim to automate such work in offices do not have those capabilities. They require precise descriptions of their function and how to perform it. Therefore, they can "commit" to doing only a formalizable approximation of the function desired by the client. They are incapable of performing the function in situations that do not match the assumptions of the formalization, and can not adapt their methods to account for unexpected features of a particular situation such as resource limitation changes or interactions with other tasks. In addition, they require more effort by the client to establish their task or function contract since they have no capability of interpreting vague descriptions and only very limited capabilities for recognizing situations where a description is inapplicable.

All too often, designers and installers of office automation equipment do not realize the unformalizable subtleties of the work being automated, and therefore do not anticipate the differences between what the equipment is going to do and what the people did whom it is replacing. Those differences often cause major upheavals in an organization because they change the work requirements of all the agents who interact with the equipment. A major goal of the analysis described in this paper has been to provide a model of the unformalizable aspects of office functions being overlooked by current automation efforts so that the differences in functionality introduced by the automation can be predicted and compensated for.

Automation can increase productivity in an office by supporting, as well as replacing, people in their performance of office functions. For example, the framework we have described suggests ways of supporting office work by providing agents with the information they need when they need it. It also suggests a facilitator role for a computer-based system using knowledge of who the clients, contractors, and consumers are for each task being performed. By knowing who must be satisfied by each result, a system would be able to monitor and track the performance of a task without needing to understand the methods being used or the semantics of the task itself.

Acknowledgements

Lucy Suchman is a major participant in the research effort from which this paper emerges and has made significant contributions to the ideas presented herein. Also, I would like to thank Danny Bobrow, Austin Henderson, Tom Malone, Al Newell, and Kurt VanLehn for providing valuable feedback and ideas in support of this paper.

References

Fikes, R. "Automating the Problem Solving in Procedural Office Work" In Proceedings of the AFIPS Office Automation Conference, Houston, Texas, March 1981.

Fikes, R., and Henderson, A. "On Supporting the Use of Procedures in Office Work" Proceedings of The First Annual National Conference on Artificial Intelligence, Stanford, Calif., August 1980.

Fikes, R., and Nilsson, N. "STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving". Artificial Intelligence 2 (1971), pp 189-208.

Flores, F., and Ludlow, J. "Doing and Speaking in the Office". In Fick, G., and Sprague, R. (eds), *DSS: Issues and Challenges.* London: Pergamon Press, 1981.

Smith, R. G. "A Framework for Problem Solving in a Distributed Processing Environment". Dept. of Computer Science, Stanford Univ., Stanford, Calif., STAN-CS-78-700 (HPP-78-28) Dec. 1978.

Smith, R. G., and Davis, R. "Frameworks for Cooperation in Distributed Problem Solving" IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-11, No. 1, January 1981.

Suchman, L. "Office Procedures as Practical Action: Theories of Work and Software Design" To appear in a special issue on Office Semantics of the IEEE Transactions on Software Engineering.

VanLehn, K., and Brown, J. S. "Planning Nets: A Representation for Formalizing Analogies and Semantic Models of Procedural Skills". In Snow, R., Frederico, P., and Montague, W. (eds) *Aptitude Learning and Instruction: Cognitive Process Analyses.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

A COGNITIVE SCIENCE APPROACH TO IMPROVING PLANNING

Barbara Hayes-Roth
Rand Corporation
Santa Monica, CA. 90406

INTRODUCTION

Planning is the predetermination of an intended sequence of actions aimed at achieving a goal. We all engage in planning for a variety of goals, ranging from everyday goals like performing a set of errands to more consequential goals like making a career change. Whether or not we achieve our goals depends in part on the quality of our plans.

During the past few years, my colleagues and I have been studying the cognitive processes people use for planning. When we began this work, most of the earlier scientific research on planning had focused on the development of automatic planning systems (e.g., Fahlman, 1974; Fikes, 1977; Sacerdoti, 1974, 1975). Other research had examined the role of plans in human behavior (e.g., Ernst & Newell, 1969; Miller, Galanter, & Pribram, 1969). However, little was known about the psychology of planning per se--how to identify effective planners, what special skills or strategies effective planners use, and what task factors impact on planning effectiveness. Because our long-range goal was to develop computer aids for human planners, we felt that understanding these psychological issues was an important first step. Accordingly, we embarked on a program of research designed to elucidate the cognitive processes underlying planning and to develop a computer aid that exploits cognitive strengths and compensates for cognitive weaknesses.

Of course, there are many different kinds of planning, depending upon the number of planners involved, the planning environment, the type of goals under consideration, the action options, etc. For our research, we chose to focus on individual planning of multiple-task sequences in a spatial environment. This task domain is well-defined and manageable. At the same time, it is general enough to apply to a variety of specific real-world planning tasks (e.g., planning travel itineraries, planning delivery routes, planning factory inspections, planning tactical missions). For our research, we wanted an instantiation of this task that was both realistic and familiar to the people who would serve as subjects in our experiments. We chose the following errand-planning task:

Given: a list of desired errands
a map of the local environment
starting and ending times
starting and ending locations
temporal constraints
contextual information

Plan: which errands to accomplish
how much time to allocate for each errand
in what order to conduct the errands
by what routes to travel between successive errands.

Our research program comprises the following tasks:

1. Develop a cognitive model of the planning process.
2. Conduct experimental investigations of the model.
3. Apply the model in studies of individual differences in planning skill.
4. Apply the model in studies of planners' deficiencies and their vulnerabilities to task factors.
5. Infer principles for improving planning performance.
6. Design a computer aid around the inferred principles.
7. Implement the computer aid.
8. Test the computer aid in real planning environments.

We have completed tasks 1-5 and have recently begun working on task 6. This paper summarizes our work to date.

A COGNITIVE MODEL OF THE PLANNING PROCESS

Our model of the cognitive processes underlying planning behavior exploits the theoretical architecture of the Hearsay-II speech-understanding system (CMU Computer Science Research Group, 1977; Erman & Lesser, 1975; Lesser, Fennel, Erman, & Reddy, 1975; Lesser & Erman, 1977; Hayes-Roth & Lesser, 1977). It also incorporates principles developed in the research on automatic planning systems and on the role of plans in human behavior mentioned above. The model has three basic components: specialists, a "blackboard," and a control regime. Each of these is discussed briefly below. (See Hayes-Roth & Hayes-Roth, 1979, for a more detailed discussion of the model.)

Specialists

Specialists are the mental processes that generate decisions for incorporation into the plan in process. For example, one specialist might generate a decision to establish a particular goal for the plan. Another might generate a decision to take a particular action toward achieving that goal.

We operationalize specialists as condition-action rules, similar to the production rules of Newell and Simon (1972). The condition part of the rule describes the circumstances under which the specialist can make a contribution to the plan. Ordinarily, the condition requires that some other planning decision has already been made. When that condition is satisfied, we say that the specialist has been "invoked." The action part of the rule describes the decision the specialist can contribute to the plan if it is "executed."

We assume that a given individual possesses many planning specialists. Some of them are generic and can

make contributions to all planning problems. Other specialists are domain-specific and can make contributions only to planning problems in their particular domains. We also assume that the many specialists an individual brings to bear on a planning problem operate independently. They do not communicate or influence one another's behavior directly. However, they can communicate and influence one another's behavior indirectly, as discussed in the next section.

The Blackboard

The blackboard is a structured internal data base in which executed specialists record their decisions. All specialists can also inspect the blackboard and respond differentially to the presence of different kinds of information. In this way, specialists indirectly communicate and influence one another's behavior.

The model partitions the blackboard into five conceptual "planes" that distinguish the different kinds of decisions we think planners make. The meta-plan plane contains decisions about how to approach the problem, what kinds of problem-solving strategies to use, what kinds of policies should guide plan development, and what kinds of criteria should be used to evaluate tentative plans. The plan abstraction plane contains decisions characterizing the kinds of actions that should be included in the plan. The knowledge base plane contains data, assumptions, and knowledge about the world that might be useful in instantiating plan abstraction decisions. The plan plane contains decisions about the plan itself. These decisions are typically instantiations of plan abstraction decisions, based on information in the knowledge base. Finally, the executive plane contains decisions about how to sequence the execution of invoked specialists during the planning process. These decisions determine the order in which decisions are generated on the other planes of the blackboard.

The model further partitions each plane of the blackboard into different "levels of abstraction." To illustrate, the plan plane has four different levels of abstraction. The outcomes level contains decisions about the goals of the plan. The design level contains decisions about the general spatial-temporal organization of the plan. The procedures level contains decisions about the actions planned within that spatial—temporal organization. The operations level contains decisions about the low-level operations necessary to carry out those actions. The other planes have similar levels of abstraction.

The blackboard structure outlined above serves two functions. First, it embodies our model of the psychological categories of planning. Thus, it distinguishes our model from other planning models and provides one basis for evaluating the model's psychological validity. Second, the blackboard structure improves computational efficiency by permitting specialists to restrict their inspection activities to those areas of the blackboard likely to contain information of interest.

Control Regime

According to the model, planning proceeds in a series of "cycles." On each cycle, many specialists may be invoked. One specialist is scheduled to execute its action next. It does so, recording its decision at an appropriate location on the blackboard. The recording of a new decision signals the beginning of the next cycle. This process repeats until the planner has developed a satisfactory plan.

The process of scheduling one of the invoked specialists on each cycle is another important feature of the model. Most previous conceptions of the planning process imposed upon it a strict, hierarchical control regime. High-level abstract decisions were made first and refined by later decisions at successively lower levels of abstraction. By contrast, our model assumes an opportunistic control regime. Specialists are scheduled and decisions generated in highly variable orders determined by competing scheduling heuristics. We have concentrated on two scheduling heuristics, focus of attention and recency. The focus of attention heuristic recommends scheduling specialists that record decisions in pre-selected areas of the blackboard. The recency heuristic recommends scheduling specialists that have been invoked recently, for example, on the last one or two cycles. (Hayes-Roth & Lesser (1977) have recommended other heuristics, such as efficiency and efficacy.) We implement these heuristics by means of specialists that record relevant decisions on the executive plane of the blackboard.

The interaction of the focus of attention and recency heuristics can manifest a variety of specific control strategies, including the hierarchical strategy mentioned above. We believe that the flexibility embodied in the opportunistic control regime is both a more accurate model of the variability we observe in human planning behavior and a more powerful approach to planning in general.

EXPERIMENTAL INVESTIGATIONS OF THE OPPORTUNISTIC MODEL

We conducted two kinds of experimental investigations of the planning model: psychological experiments and computer simulation experiments.

The psychological experiments provided support for several of the basic assumptions of the model, including the following: (a) that people make planning

decisions in each of the conceptual categories of the blackboard; (b) that people formulate plans at the postulated levels of abstraction; (c) that people develop plans with both top-down and bottom-up decision sequences; (d) that people effect alternative control strategies for planning; and (e) that people opportunistically exploit the information and constraints available during planning. (These experiments are discussed in detail in several reports: Goldin & Hayes-Roth, 1980; Hayes-Roth, 1980;Hayes-Roth & Hayes-Roth, 1979; Hayes-Roth & Thorndyke, 1980.)

The computer simulation was a LISP implementation of the model with about fifty specialists. The simulation served two functions. First, it demonstrated the sufficiency of the model to account for human planning behavior. The simulation produced plans and planning protocols similar to those produced by human planners. It also exhibited their characteristic strategic flexibility. Second, the simulation allowed us to explore some of the computational properties of the model, providing more general insights into distributed computation and heuristic control regimes. (The computer simulation is discussed in more detail in Hayes-Roth & Hayes-Roth, 1979, and in Hayes-Roth, Hayes-Roth, Rosenschein, & Cammarata, 1979).

INDIVIDUAL DIFFERENCES IN PLANNING SKILL

The first application of the opportunistic model was to individual differences in planning skill--why are some planners more effective than others? The model suggests three areas in which effective planners might differ from ineffective planners: their generation of decisions in different areas of the planning blackboard, their flexibility in distributing attention among the different areas of the blackboard, and their repertoires of specialists.

We evaluated these hypotheses by examining the planning processes of several planners with different skill levels. We assessed a planner's skill level based on the quality of the plans he or she produced. The quality of a plan was a composite score reflecting several interacting dimensions (e.g., efficiency, constraint satisfaction, temporal realism). Planners who achieved high plan scores were designated effective planners; those who achieved low plan scores were designated ineffective planners. We then examined the planning process of effective versus ineffective planners as revealed in thinking aloud protocols. Basically, we asked subjects to verbalize their thoughts while they formulated plans. We then analyzed these protocols, classifying statements as representing particular planes of the blackboard or levels of abstraction. Finally, we examined the relationship between planning skill and the frequency with which planners made different kinds of decisions.

The results supported all three hypotheses advanced above. Effective planners generated decisions in all areas of the planning blackboard, whereas ineffective planners generated primarily plan and plan-abstraction decisions. Effective planners also generated decisions at different levels of abstraction, whereas ineffective planners generated primarily low-level decisions. Effective planners showed greater attentional flexibility than ineffective planners. They more frequently shifted their focus of attention among the different planes of the blackboard and among different temporal loci within the plan itself. Finally, effective planners exhibited many more specialists than ineffective planners and they seemed to exploit powerful specialists more actively. (These results are discussed in more detail in Goldin & Hayes-Roth, 1980.)

PLANNERS' DEFICIENCIES AND THEIR VULNERABILITIES TO TASK FACTORS

We next applied the model to an analysis of general deficiencies in human planning and to the deleterious effects of task factors. Using protocol analysis methods similar to those described above, we identified three task factors that impede effective planning: constraints, complexity, and time stress.

Planners seem able to accommodate one or two simple time constraints. However, as the number of time constraints in a problem increases, planning effectiveness deteriorates. Time-constrained tasks, particularly those that appear late in the plan, rarely satisfy their constraints. The problem lies in planning strategy. The opportunistic model permits alternative planning strategies and, for heavily constrained problems, a constraint-based strategy is appropriate. Apparently, however, most of our subjects did not have a constraint-based strategy in their repertoires.

Planners also have difficulty coping with increasing problem complexity. As the number of tasks under consideration and the number of alternative possible plans increase, planners require disproportionately longer times to generate satisfactory plans. This problem also seems to lie in planning strategy. Instead of adopting a strategy which would restrict attention to a small number of the most promising alternatives, our subjects appeared willing to consider new alternatives throughout a planning session.

The third deficiency in human planning lies in the area of time estimation. Planners tend to underestimate the time required to execute planned actions and, as a consequence, to overestimate the number of actions they can execute in the available time. This tendency is exacerbated by time stress (the

time required to execute all the actions under consideration divided by the available time). Two factors contribute to this problem. Planners typically estimate time requirements at relatively high levels of abstraction. Because they fail to enumerate all time-consuming components of an action, they systematically underestimate the time required to execute it. Planners also respond to strong motivational factors. Their strong desire to accomplish all or most of the tasks under consideration biases their time estimates toward underestimation. (These results are discussed in more detail in Hayes-Roth, 1980.)

PRINCIPLES FOR IMPROVING PLANNING PERFORMANCE

Based on the opportunistic model and the empirical results summarized above, we developed the following principles for improving planning performance:

Criteria for Selecting Planners

1. Large-capacity working memory. The opportunistic model describes people's tendency to "jump around" the space of possible considerations while forming plans. This suggests that it may be important for planners to have large-capacity working memories in order to keep track of several aspects of a developing plan simultaneously.

2. Attentional flexibility. Our studies of individual differences in planning performance showed that good planners shift attention among different aspects of a planning problem more frequently than poor planners. Therefore, attentional flexibility may be another important characteristic of potential planners.

3. Strategic flexibility. Our studies of top-down versus bottom-up planning strategies showed the impact of particular planning strategies on the efficacy of the plans subjects produce and on the efficiency with which they produce them. In addition, some subjects appeared more willing than others to adopt alternative strategies. Therefore, strategic flexibility may be another important selection criterion.

What to Teach Planners

4. Concepts of abstract plans, meta-cognitive decisions, executive decisions, and knowledge-base decisions. Our studies of individual differences showed that good planners made decisions in all categories of the planning blackboard, whereas poor planners made only certain kinds of decisions. In particular, high-level abstract decisions, world knowledge decisions, metacognitive decisions, and executive decisions distinguished good planners from poor planners. Therefore, planners should be taught the roles of these types of decisions in the planning process.

5. Domain-specific planning heuristics. Our studies of individual differences also showed that good planners had more different planning specialists than poor planners. Therefore, planners should be taught a variety of domain-specific planning heuristics.

6. Costs and benefits of opportunism. There is considerable evidence that most people employ some amount of opportunism in the planning process. Opportunism provides planners freedom to examine various aspects of a problem, investigate alternative plan configurations, etc. This enables them to discover solutions that more rigid approaches would obscure. On the other hand, opportunism requires additional time and may lead planners down unproductive, as well as productive, solution paths. Planners should be taught these advantages and disadvantages and how to exercise controlled opportunism.

7. General planning strategies. As mentioned above, different planning strategies are appropriate under different circumstances. Planners should be taught general planning strategies and the circumstances under which each is appropriate.

8. Judgment and time estimation. Most people show a strong tendency to underestimate the time required for planned actions. As a consequence, their plans are unrealistic and overrun the time available for execution. Planners should be taught cognitive methods for making such judgments more reliably and more accurately.

How to Train Planners

9. Provide explicit instruction. Explicit instruction appears to be a highly effective technique for training particular planning strategies and methods.

10. Induce illustrative experiences. Many planners seem to be able to generalize what they learn from one planning problem to subsequent, similar planning problems. Therefore, general strategies and methods can be taught by instructing planners how to use them on specific problems and providing opportunities for them to generalize them to similar problems.

11. Illustrate effective planning. Because our planning simulation effectively mimics the cognitive processes people use while planning, it may provide a useful model during training. The simulation could be programmed to illustrate planning strategies.

Useful Aids for Planners

(See the following section.)

ONGOING RESEARCH

As discussed in the introduction to this paper, we are particularly interested in developing computer aids to planning. We believe that, for the foreseeable future, people will play central roles in most important planning activities. Accordingly, an effective planning aid should exploit people's cognitive strengths and compensate for their cognitive weaknesses. We have recently begun working on the design of such an aid. Our work in this area focuses on a different instantiation of the same general planning task: project planning. We chose this task because it is an important real-world task that could benefit from the development of a planning aid and because we have contact with a variety of people who carry on project planning professionally. Our current design comprises the following components:

Goal Formulation

Our research suggests that planners suffer three deficiencies in the area of goal formulation. They do not formulate complete, well defined goals specifying all of the objectives, constraints, and policies the plan in progress should serve. They sometimes vacillate among conflicting goals, striving to satisfy different goals at different points in the planning process. They try to satisfy too many goals, given the available resources. The goal formulation component should help planners to articulate, prioritize, prune, and coordinate project goals.

Product Specification

Given a set of goals, the planner must generate functional specifications for project product(s). Presumably, development of these products would satisfy project goals. The problems in product specification are that planners may not generate complete specifications or they may not specify products that systematically satisfy project goals. The product specification component should help them to do so.

Task Analysis

Given a set of specifications, the planner must specify a set of project tasks whose execution would implement planned project products. Again, the problems are that planners may not generate a complete set of tasks or analyze them at a sufficiently low level of abstraction or that they may not specify tasks that systematically implement project products. The task analysis component should help them to do so.

Resource Estimation

For each task under consideration, the planner must determine what resources are required to execute it. Our research suggests that planners are unduly optimistic about the number of tasks that can be accomplished with given fixed resources. The resources estimation component should help planners realistically assess the resource requirements of tasks under consideration.

Resource Allocation

Given limited resources and alternative goals, the planner must determine how to allocate the available resources. Our research suggests that planners tend to spread resources too thinly across too many goals. The resource allocation component should help planners to formulate realistic allocation schemes and to perform cost/benefit analyses of alternative allocation schemes.

Scheduling

The planner must schedule planned tasks in a way that provides adequate time for the execution of each task, insures completion of prerequisites by the time they are required, and minimizes slack time and other costs. The scheduling component should support these activities.

Evaluation

Our research suggests that poor plan evaluation is a major impediment to effective planning. Poor planners do very little systematic evaluation of their tentative plans. Even good planners frequently decide arbitrarily among two or three final contenders. The evaluation component should at least assess whether the final plan meets criteria articulated by the planner. It should also assess the efficacy, efficiency and robustness of the plan in a simulated environment.

During the next few years, we plan to refine this design, implement it as a computer system, and evaluate its utility in the context of several Rand projects.

REFERENCES

CMU Computer Science Research Group. Summary of the CMU Five-year ARPA Effort in Speech Understandng Research, Technical Report, Carnegie-Mellon University, 1977.

Erman, L.D. and V.R. Lesser. "A Multi-Level Organization for Problem Solving Using Many Diverse Cooperating Sources of Knowledge," Proceedings of the Fourth International Joint Conference on

Artificial Intelligence, Tbilisi, USSR, 1975, 483-490.

Ernst, G.W. and A. Newell. GPS: A Case Study in Generality and Problem Solving, New York: Academic Press, 1969.

Fahlman, S.E. "A Planning System for Robot Construction Tasks," Artificial Intelligence, 1974, 5, 1-49.

Fikes, R.E. "Knowledge Representation in Automatic Planning Systems," In A.K. Jones (ed.), Perspectives on Computer Science. New York: Academic Press, 1977.

Goldin, Sarah and Barbara Hayes-Roth. Individual Differences in Planning Processes, N-1488-ONR, The Rand Corporation, June, 1980.

Hayes-Roth, Barbara and Frederick Hayes-Roth. "A Cognitive Model of Planning," Cognitive Science, 1979, 3, 275-310.

Hayes-Roth, Barbara and Frederick Hayes-Roth. Cognitive Processes in Planning, R-2366, The Rand Corporation, December, 1978.

Hayes-Roth, Barbara and Perry W. Thorndyke. Decisionmaking During the Planning Process, N-1213-ONR, The Rand Corporation, October, 1980.

Hayes-Roth, Barbara. Estimation of Time Requirements During Planning: Interactions Between Motivation and Cognition, N-1581-ONR, The Rand Corporation, November, 1980.

Hayes-Roth, Barbara. Flexibility in Executive Strategies, N-1170-ONR, The Rand Corporation, September, 1980.

Hayes-Roth, Barbara. Human Planning Processes, R-2670-ONR, Rand Corporation, December, 1980.

Hayes-Roth, F. and V.R. Lesser. "Focus of Attention in the Hearsay-II Speech Understanding System," Proceedings of the Fifth International Joint Conference on Artificial Intelligence, Boston, Mass., 1977, 27-35.

Lesser, V.R., R.D. Fennell, L.D. Erman, and D.R. Reddy. "Organization of the Hearsay-II Speech Understanding System," IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-23,1975,11-23.

Miller, G.A., E. Galanter, and K.H. Pribram, Plans and the Structure of Behavior, Holt, Rinehart and Winston, Inc., 1960.

Newell, A. and H.A. Simon. Human Problem Solving. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Sacerdoti, E.D. "Planning in a Hierarchy of Abstraction Spaces," Artificial Intelligence, 1974, 5, 115-135.

Sacerdoti, E.D. A Structure for Plans and Behavior. Technical Note 109, Stanford Research Institute, Menlo Park, California, August, 1975.

Everyday Problem Solving

James A. Levin

Laboratory of Comparative Human Cognition
University of California, San Diego

This paper was generated through the distributed social processing of the Laboratory of Comparative Human Cognition, with special help from Denis Newman, Naomi Miyake, Bud Mehan, Ed Hutchins, Peg Griffin, Mike Cole, and Marcia Boruta. Financial support was provided by The Spencer Foundation.

Abstract

Everyday problem solving is different in significant ways from the kinds of problem solving that take place in laboratory microworld settings. Attempts to simplify have excluded important factors that can help us understand aspects of the problem solving that are problems from the point of view of the laboratory. This paper describes several research projects that have examined problem solving in non-laboratory settings, and some of the implications of these studies for cognitive science. The current notions of distributed cognitive processing can be extended in a powerful way to the socially distributed problem solving characteristic of everyday settings. This notion of socially distributed problem solving can then reflect back on individual problem solving, which which is acquired and often carried out in social settings.

Someone walks into your office and asks you to recommend a paper to read as an introduction to research on children's problem solving. You discuss with the person exactly what she wants to know, you walk over to your bookshelf to look for an appropriate book, you call a friend on the phone who might know. All very unexceptional, yet imagine that the person didn't allow you to discuss with her exactly what she wanted to know, to go to your bookshelf or call on the phone, but instead required you to answer her question without these external resources. In everyday circumstances you would throw her out of your office. Yet these are exactly the constraints of the laboratory microworlds within which problem solving is studied.

Recently a number of research groups have been studying problem solving in non-laboratory settings. This work has some important implications for cognitive science: it serves to reinforce the findings concerning the role of expertise in human problem solving and expert artificial intelligence systems, and this non-laboratory work extends beyond the current problem solving research, pointing to ways to enrich both models of human problem solving and expert inference systems.

Expertise.

Recent research on expert problem solving has highlighted the large amount of domain specific knowledge and cognitive processes that constitute expertise (Chase & Simon, 1973; Larkin, McDermott, Simon, & Simon, 1980). These studies of human problem solving have been parallelled by the development of artificial intelligence "expert" systems, which are also characterized by a focus on domain specific knowledge and inference processes.

This work contrasts with the early work on problem solving, both in psychology and in computer science, which postulated a few all powerful general problem solving processes, that would operate over a large passive data base. These "central processor" models have been displaced by various "distributed" data base and processor models, with multiple concurrent processes that interact to produce complex processing.

The domain specific focus of expert knowledge has been reinforced by studies of everyday problem solving. For example, a group of researchers at the University of California, Irvine have been examining the ways that ordinary adults use arithmetic knowledge while grocery shopping (Lave, 1980; De la Rocha, Murtaugh, & Lave, 1981). Schools spend many years teaching us general purpose algorithms that can be used during shopping to calculate comparative prices. Yet most of their observations show that people ordinarily use special purpose heuristics while shopping. Even in this mundane everyday setting, people have developed "expertise" to carry out this task, domain specific methods that bear little resemblance to the general purpose computational skills taught in school.

Similar research by Scribner and her associates (Scribner, 1981) reinforce this finding of special purpose expertise in everyday functioning. They studied the work in a dairy warehouse, examining the use of computation in filling orders and determining total prices. The experts in this domain had developed special purpose algorithms to allow them to function efficiently in this domain.

Problem Solving vs. Routine Functioning.

What is the relation between expertise and problem solving? Problem solving is not just accomplishing particular kinds of tasks labeled as "problems". The processes involved in solving most puzzles are different the second time you solve them (when you know the answer) than the first time. In fact, "expertise" can be defined as the knowledge and cognitive skills that allow a person to perform routinely what other people would have to do through problem solving. Central to the definition of problem solving is the notion of a "blocked condition" (Hutchins & Levin, 1981). Derived from the Gestalt studies of problem solving (Kohler, 1925), this occurs when a problem solver is unable to achieve some goal, after repeated attempts to do so. Problem solving is the cognitive processing that occurs when a problem solver is blocked. Routine functioning is the processing that occurs when unblocked.

Studies of everyday problem solving.

Several research groups have examined how people deal with these blocked conditions in non-laboratory settings. Suchman (1980) did an ethnographic study of problem solving in an office setting, setting down a detailed account of accounting practices, especially those involved in dealing with non-standard cases. Even in the mundane work-a-day office setting, the execution of explicit instructions remains "irremediably problematic", requiring interactive work on the part of the participants. Levin & Kareev (1980) examined the problem solving of children in computer clubs. In both these studies, a critical component of the problem solving, which is largely absent from laboratory studies, is the conceptual organization of the task, determining what the problem IS, what are the goals and constraints to be satisfied, what actions are available.

The second major difference between laboratory problem solving and everyday problem solving is the much more important role played by external resources, those outside the individual problem solver. The laboratory setting is relatively sterile of help and the experimenter works to keep it that way. A standard experiment would not be run in the middle of a busy room. Given a puzzle to solve, it is not considered proper for the subject to ask a friend what the answer is. However, when a person encounters a problem in everyday life, asking someone what to do is usually appropriate.

The use of social resources is probably the biggest single difference between standard laboratory settings and everyday settings. One of the common strategies used by adults when faced with an

arithmetic problem in an eveyday setting is to ask someone what the answer is (Lave, 1980). In computer clubs, children help and ask for help effortlessly to get beyond minor blocks so that they can get on with their play (Levin & Kareev, 1980).

Division of labor. One of the most important ways that people use social resources for problem solving is by organizing a division of the work involved. When faced with a new computer game, children divide up the task so that no child is overwhelmed while the game gets played. One child will type in the required responses, another will take over the generation of guesses, others will evaluate and modify the guesses (Levin & Kareev, 1980). This process of organizing and executing a division of labor is so effortless and smooth that repeated viewing of video taped instances is required to see it at all.

Socially distributed problem solving.

An important contribution of cognitive science to the study of everyday problem solving are the frameworks for distributed processing. Using this new processing "language", researchers can now talk about the social distribution of problem solving in many situations, characterizing the nature of each processor and the kinds of interactions that occur (LCHC, 1981; Mehan, 1981). Issues of conflict resolution and information integration, central issues for distributed processing, are also critically important for models of socially distributed problem solving.

Yet this contribution is not a one-way street. The study of problem solving that is distributed over several people can suggest hypotheses about how the same process might be organized cognitively when performed by a single person. A major issue for any cognitive model is what is the structure of the knowledge and processes, what are the units and subunits. The division of a problem that can be solved distributed across several people provides an existence proof that the problem can successfully be organized that way. Researchers can then carry out empirical tests of whether an individual in fact does organize the task that way.

Aquisition of expertise. A second contribution of this approach to everyday problem solving is to deal at least partly with the issue of acquisition. Current research on the acquisition of problem solving skills and domain specific expertise has concentrated on independent invention (Langley, 1980; Lenat, 1977). Yet models that depend totally on independent invention of knowledge and processes never get very far. It remains a major puzzle how such systems could acquire the huge amounts of domain specific knowledge needed by experts. A way to overcome this block was pointed out by D'Andrade, in his invited address at the previous Cognitive Sciences Meeting (1980): people acquire knowledge and skills from other people. We are socialized within a rich culture, where the people and objects are at least partially organized specifically to help novices become experts in the domains important for functioning in the world. Children are taught the important facts of life; beginners are trained to become experts.

From this point of view, the acquisition of expertise can be characterized as the progressive internalization by the learner of socially distributed processing. A person generally becomes an expert in a setting where he/she can gradually take on more and more of the effort in handling tasks that experts in the domain handle. Children who are novices at a computer game initially divide the task over several children and adults. Gradually, each child takes over more and more functions, so that fewer children have to cooperate to accomplish the task. Finally a child can play the game alone. This progressive acquistion of a process initially distributed socially in fact can provide a rationale for the way that the cognitive processes are distributed within an individual expert, a distribution that allows the expert to draw smoothly upon social resources whenever problems arise.

References

Chase, W. G., & Simon, H. A. Perception in chess. Cognitive Psychology, 1973, 4, 55-81.

De la Rocha, O., Murtaugh, M., & Lave, J. A conceptual framework for locating cognitive processes in everyday life. To appear in J. Lave & B. Rogoff (Eds.), Everyday cognition: Its development in social context. Cambridge, MA: Harvard University Press, in press.

D'Andrade, R. G. The cultural part of cognition. Paper presented at the Second Annual Cognitive Science Conference, Yale University, 1980. Cognitive Science, in press 1981.

Hutchins, E. L., & Levin, J. A. Point of view in problem solving. Paper presented at the Third Annual Cognitive Science Conference, U. C. Berkeley, 1981.

Kohler, W. The mentality of apes. London: Routledge and Kegan Paul, 1925.

Langley, P. Data-driven discovery of physical laws. Cognitive Science, 1981, 5, 31-54.

Lave, J. What's special about experiments as contexts for thinking. The Quarterly Newsletter of the Laboratory of Comparative Human Cognition, 1980, 2, 86-91.

Laboratory of Comparative Human Cognition, UCSD. Culture and intelligence. In R. Sternberg (Ed.), Handbook of intelligence. New York: Cambridge University Press, 1981.

Larkin, J., McDermott, J., Simon, D.P., & Simon, H.A. Expert and novice performance in solving physics problems. Science, 1980, 208, 1335-1342.

Lenat, D. B. Automated theory formation in mathematics. Proceedings of the Fifth International Joint Conference on Artificial Intelligence, 1977, 833-842.

Levin, J. A., & Kareev, Y. Problem solving in everyday situations. The Quarterly Newsletter of the Laboratory of Comparative Human Cognition, 1980, 2, 47-52.

Mehan, H. Practical decision making in naturally occurring institutional settings. To appear in B. Rogoff & J. Lave (Eds.), Everyday cognition: Its development in social context. Cambridge, MA: Harvard University Press, in press.

Newell, A., & Simon, H. A. Human problem solving. Englewood Cliffs, NJ: Prentice-Hall, 1972.

Scribner, S. Studying working intelligence. To appear in J. Lave & B. Rogoff (Eds.), Everyday cognition: Its development in social context. Cambridge, MA: Harvard University Press, in press.

Suchman, L. Office procedure as practical action: Theories of work and software design. Paper presented at the Workshop on Research in Office Semantics, Chatham, Massachusetts, June 1980.

Marriage is a Do-It-Yourself Project: The Organization of Marital Goals

Naomi Quinn
Duke University

In the story understanding literature, a class of stories which is on its way to becoming something of a famous example, has to do with getting married. One typical variant, used by Wilensky (1978: 67), goes like this:

> John was tired of frequenting the local singles' bar. He decided to get married.

The point of this story is that to understand how getting married could possibly replace going to a singles' bar, the reader must know that a state such as marriage can subsume goals that arise repeatedly. In this case the goal is Satisfy-SEX. A series of variants of this story, however, demonstrate that marriage may in fact subsume a variety of such recurring goals:

> John was feeling lonely every evening. He decided to get married (Wilensky ibid.: 72).

> Mary was tired of working for a living. She decided to quit her job and get married (Wilensky ibid.: 77).

Schank and Abelson introduce two other variants (1977: 126) to illustrate a further characteristic of goal subsumption:

> After his marriage with Mary broke up, John began frequenting the local singles' bars.

> After his marriage with Mary broke up, John decided to join a chess club.

To understand these stories we need to realize, Schank and Abelson (ibid.) say, that "goals that arise via subsumption tend to subsume more than a single goal." "Being married," they note, "can subsume the sex urge goal, goals of social stimulation, the desire to have children, to have power over another person, or to be with a loved one." Wilensky (ibid.: 77) views this tendency to goal multiplicity as a distinctive property of social relationships:

> Since there is typically more than one obligation imposed upon the member of a relationship, social relationships usually subsume more than one goal. For example, being married to someone usually subsumes a number of recurring goals in addition to those for which money is instrumental. Since being married obligates each partner to have sex with the other, and since having a willing partner is a precondition for having sex, then marriage subsumes the recurring Satisfy-sex goal. Also, since marriage requires the partners to live together, then it subsumes the recurring Enjoy-company goal to which being near a loved one is instrumental.

These examples promote a certain view of social relationships as bundles of discrete goals which happen, perhaps because of cultural convention, to be packaged together. As Schank and Abelson (ibid.: 127) summarize, "Each social relationship carries with it a packet of goals." This paper is offered as an alternative to that particular view of goals and social relationships. It argues that marital goals are not so much packaged together as highly organized, and that what organizes them is an underlying model of marriage which, far from being fixed and conventional, is constructed by the individual in the course of the relationship. Perhaps the best way to introduce this view of the organization of goals is to quote some statements by Alex, an interviewee of mine, on the subject of his marriage. Talking about his preconceptions of marriage, he says,

> 6H-1: Well yeah I thought it was all going to be wonderful. You know it was--the problem of sex was going to be solved. You know I was an adolescent or barely out of adolescence, you know--this was a wonderful idea. And, I don't know...you know--the idea--and--you really--you're asking some good questions because there really are some things that I knew about and that I wanted. A companion and friend. Probably the most. You know, the things that I did want and I got out of the marriage. That seemed important then--to have someone there all the time that you could rely on. And talk to all the time about things. Somebody to help and somebody to help you, you know, that seemed like a real good idea. That seemed like something I really wanted. It seemed like something that we got out of the marriage. Somebody always there.

In a second interview, he goes back to this issue:

> 6H-2: I think, right from the very beginning, I believe I mentioned on the other tape that I felt the--talking about the--what was--did you expect getting married that I did feel a need for a companion and somebody to be with me and to fulfill something that was empty, I didn't know what it was. And I think that might have been the first step--the first stage of love. But now I'd say that love is, to me, is the desire to give more to the other person than you're giving to yourself, at times. You know it's the, not only willingness to accept the part of marriage where you have to change and adjust, but a strong desire to do so. Not only the willingness to accept that there are times when the other part--person or partner might need some help. But the desire to give help. To that person. The sort of love of the fulfilling the other person's needs.

Later in the same interview, talking about why some of the couples they had been thrown together with in the Navy had marital difficulties:

> 6H-2: There were certain things they married for. They'd gotten which it may have been their dinner on the table every night and, you know, a warm body next to you in bed. And it suddenly went beyond that. It suddenly got into conversation. And support, things of that sort. And that's--that caused some difficulties.

In a later interview, he describes giving advice to a Navy shipmate:

> 6H-4: And he said he was going to get married and I said, "Well," I said, "I hope you think about it real hard because I think you might find marriage to be a little bit surprising than what it is. Because it was for me. Shocking sometimes, you know, that it wasn't all love and sex and that's it. Yeah, that there was some work to be done."

The first thing to be said about the "goal subsumption" model of marriage, then, is that it is a folk model. The reason why we understand marriage as subsumption of a particular set of goals, is not because social relationships necessarily subsume multiple goals, but because the goal subsumption model of marriage is a very common preconception in our society. It is not,

however, the only myth about marriage; another, sub-cultural one, is expressed in the following segments from interviews with another husband,

> 2H-7: The marriage thing, I think basically, is just a mental thing where you know, you've heard so many things about marriage like it's--marriage have become something like a negative thing. You know, a lot of people, you know, you know, it's just--you just hear so much, you've never been married so you don't know. You know you just hear so many things about when you're married, you know, things really change. Your wife don't let you go out or, you know, your husband don't let you go out and, you have to--you know, you never have any money, you can't spend money on what you want to. You know you hear so many things--when you get married your wife don't want you sexually anymore or you don't want your wife and you have to be sneaking here and there trying to do things. You know, these are things that you hear and because of all this, you know, you know, it's like you have a--it's just like you sort of take caution and say, "Well let's try it out first and then we'll"--you know--
>
> . . . . . .
>
> You hear it all your life from a kid on up and from various people.

> 2H-7: Your wife won't let you out and--or you got to be doing something. You got to keep the baby, you're babysitting, you know. And they say things like, "You're babysitting while your wife is out in the streets, you know, playing around." Or you know, you know, they put a lot of, you know, it's a lot of fears being--it, you know, it be a lot of fears--a lot of things that people--you know, it's just constantly, you know.

> 2H-7: You know, so it's just something constantly you know, it's like something--like it's the end of the road, you know, to get married. It's like you--you're just crazy. You're out of your mind, you know.

> 2H-7: Okay, I'm single, okay and I'm going out and maybe one of my--somebody I know a buddy or somebody that's married. Okay, you hear from women and men. Okay, and you're single and people tell you, "Oh, you're so lucky." You know, you know, "You're single and you can go out here and see all these women you want to see, you know, anywhere--you know, you're just--you're living the life, you know, and they're making you feel like, "God," you know, "Marriage is like the end of everything," you know.

Marriage, in this view of Jimmy's does not so much solve the problem of sex as create it:

> 2H-7: Marriage or whatever it is--it's a sex thing, you know. You got to--your sex got to be cooking all the time, you know, or else, you know, you ain't *doing* nothing, you know. And you're wife may essential--or whoever may just leave you because, you know--or you may have to leave, you know, your wife or your woman because--or go for somebody that's, you know, that things are rolling, you know. This is what society has put into you.

Because of this very strong stereotype he has been taught, Jimmy decided he is never going to get married,

> 2H-2: I wouldn't want anything like a marriage to interfere with my normal way of living. I wouldn't want to interrupt my life to be--you know, maybe because I feel like I would probably be unhappy because no woman would really give me, maybe, the freedom that I need.

Jimmy's model might be dubbed the "goal-frustration" model of marriage. Our cultural knowledge of this model should assist us in understanding stories such as,

> Jimmy wanted his freedom. He decided not to get married.

or

> Jimmy did not want to feel like he had to keep sex cooking all the time. He determined never to get married.

The real-life excerpts above are taken from interviews with two husbands in a larger study, during which each spouse in eleven marriages has been interviewed for an average of 15-16 hours apiece. Interviewees describe their preconceptions of marriage, and make clear in the course of these extended interviews what they think their marriages turned out to be. To illustrate how goals are organized by these understandings of marriage which emerge in the course of it, I would like to provide a somewhat extended analysis of Alex's conceptualization of his marriage, which I will then compare with a somewhat briefer look at another marriage.

First of all, Alex sees his marriage as a joining together of two people, expressed more specifically in two metaphors he uses again and again: MARRIAGE IS BEING A COUPLE and MARRIAGE IS A PERSON. The first of these metaphors is characteristic of his descriptions of the early period of his thirteen-year marriage, for example in this remembrance of the decision to have a child while he was in the Navy:

> 6H-1: After I got the first promotion with the idea that the second one might be coming. I think that's how it was and that was quite successful because Shirley got pregnant right away. And she, you know, when I got back from Guantanamo she told me and we told the news to everybody and it was a really big deal. And I think that the apartment and the baby and all of that stuff really began to come down on us, you know, and we started believing that we were truly a couple. And we were truly a family and really married.

Here the equation between being married and being a couple is explicit; and being a couple is also equated with being a family. All three continue to be used interchangeably throughout Alex's interviews. The first identity is explored in the following remarks about a couple they knew in the Navy:

> 6H-2: You know it's funny there's even a couple that we got to know in Iceland that were just dating there, they just met there. One of the teachers and a fellow that was up there who didn't get married for years, but did later. I mean they went off on their own--he went further--he went out to the West Coast and got a ship. She went to Okinawa to teach. And they eventually got married. And we knew them very well, and were close to them and they were just dating then. And they ended up getting married and having a very solid marriage, now with a child. And we see them, they lived in Okinawa for a while and then they went to Germany. And that's another--it's amazing how that happened.

I: Why?

6H-2: Well that--I still feel like--I still--I knew them as a couple, even then. You know I always saw them as a couple, even though they really--they weren't in the same relationship as the other couples that we knew, but they ended up that way. What's amazing is how that couple formed the same sort of marriage that the other married couples that we knew already had.

Alex is amazed that his perception of them as a married couple turned out to be accurate. The second equation, between being a couple and being a family is elaborated further in this segment:

6H-1: That, you know, we were now a family. Just the two of us, we were a family, and that--and the commitment was to make that family move. You know, that it was going to be a solid thing--you know it was going to be a--you know just like the parents' had been--they were--and all the people we knew, around us. Remember of course that this mass divorce thing which is going on now, hadn't--was not going on then. This was--we got married in '67, I guess it was. January '67.

A couple is a family. This excerpt also introduces a goal--to make that family move. The juxtaposition of this observation with the comments about contrasting marriages back then with the mass divorce of the present time makes clear that what Alex means by a family that moves is a marriage which does not end in divorce. Thus one goal of BEING A COUPLE (and hence a family) is permanence. This excerpt also hints at how that permanence is to be attained--an issue which will be taken up below.

The COUPLE metaphor is associated not only with the goal of permanence, but also with the further goals (Preservation Goals, in Schank and Abelson's ibid.: 115-116) of exclusivity, proximity, and shared experience. The first of these goals emerges in Alex's descriptions of several incidents early in the marriage, revolving around his wife's flirtatiousness and untoward interest in men. Alex has this to say:

6H-1: I think I decided it was time for Shirley to stop fooling around with any other guys. And that [GETTING MARRIED] was the one way I thought I could convince her to do that. Other than that she was going to party for the rest of her life.

from which it is clear that being married implies exclusivity. The relationship between BEING A COUPLE and exclusivity is still clearer in another striking passage about his response to her emotional tie to her parents:

6H-1: I think it was significant because you know I made a very strong decision then and, you know, we--Shirley went along and I think that was one of the first times that I had said, "I'm going to separate you from your parents." In a very decisive manner and "This is what we're going to do." And you know, "You're going to follow me here, you're going to go with me here as your husband." Maybe it was, you know--looking back on that I--maybe that seems--sounds a little chauvinistic but I think it was a declaration that we have a marriage and it's just two of us in this marriage not four.

This segment offers an unusually clear glimpse into the internal logic by which a goal follows from a metaphor--probably because Alex is here recapping the very argument which he used to persuade his wife. Thus "we have a marriage" sets out the initial assumption; "it's just the two of us in this marriage" establishes the identity of a "marriage" with a "couple"; "not four" is a logical inference following from "just the two of us," and in turn supplies a reason for the goal of exclusivity, here to be realized by separating her from her parents. Such a chain or reasoning is not often made explicit in these descriptions of marriages; rather, the link between a given marital metaphor and a given marital goal must often be inferred from their juxtaposition, from more scattered pieces of logic which can be pieced together to relate them, and from their recurrence and emphasis as common themes in a given individual's interviews. Thus, it does not surprise us to learn that Alex repeatedly stresses the further goal of proximity. Talking about their decision whether or not to accept a post in Iceland, a station which provided housing for families, he says:

6H-2: I think if you're separated by necessity, if there is no choice about it, that's one thing. Then you become--you know, you're tied together by your letters and you're both fighting it. And that's a different thing.

. . . . . .

But to be given a chance--to be told that you may stay together that everything had been cleared, that we had a place, it had all been settled, and then to say, "I'm not going to go." That would have been a breach of the faith of the marriage. It would have been hard to overcome.

Choosing to stay together at every opportunity is a consequence of the proximity goal; and maintaining ties during forced separations is an attempt to overcome the circumstances which prevent proximity. Letter-writing was a major shipboard activity during his overseas trips in the Navy; and nowadays, on business trips, he makes long daily phone calls home. While separations can be "fought" in these ways, a marriage may not be able to sustain extremely lengthy ones:

6H-2: But I thought that would be two years like that [APART] would have done severe damage to our marriage. I think it would, you know, it would have done severe damage and if we weren't the same people that we are right now that might have happened. And that there is one of the weakening things of a trip overseas like that.

It can be inferred that separation causes damage by blocking the proximity goal, just as Shirley's flirtations blocked the exclusivity goal. Elsewhere he describes her flirtatiousness as "jeopardizing our relationship" before they were married.

Another consequence of the proximity goal is that partings are difficult; a number of incidents have to do with sorrowful departures:

6H-6: I think the thing that bothers me the most and sometimes it bothers me when I leave in the morning is if something happens to either one of us and we'll don't see each other again. And that makes me--I--every once in a while that thought strikes me--maybe it doesn't other people who are married--and it makes me very sad. And I just want to go back and say some more--and sometimes we're--it's very hard for us, you know, we'll be very slow

at parting and Shirley will some mornings, you know, just going to work Shirley'll be at the door waving to me and stuff. That's important I think to both of us that we do care enough that--and even little departures are important.

Talking about his out-of-town business trips, which he does not enjoy, Alex dwells on another goal--that of sharing experience. This can be seen as a logical consequence of the proximity goal. Shared experience both requires proximity, and affirms it.

6H-6: Now I would not choose to take a vacation alone. That would be terrible to me. I would prefer to have Shirley with me. On any sort of vacation or pleasant business at all. Any sort of real pleasant time I prefer to have her with me because if--I really would prefer to share those experiences with her. In fact, one of the things that I don't like about the trips is--those trips which are clearly just for me. When I do go to someplace very good to eat or I get to go to a show, I really feel badly about her not being there because we don't have that as a shared experience anymore. And those shows and good meals and things like that that we've done as a shared thing are really important to us and have been--are good moments for us in our marriage.

So much so that,

6H-6: I'd probably feel more guilty about going to a show than almost anything else we'd do because the two of us love to go to a show. And I would probably--I'm going to be in New York for about a week--or in Stamford for about a week in January. And I'm going to be in--about 8 days I'm going to be away in January including 4 or 5 in Stamford. And when we're in Stamford we often get tickets to New York shows--take a train in. And I'm sure we'll go see some shows. I'm going to try my best to go--to talk people into shows that Shirley would not want to go to, but I would.

And as the following excerpt shows, this goal of sharing experiences has a not inconsequential influence on the couple's economic decisions:

6H-6: But I also think that it's excellent when I--you know, when I can have the chance to have J. along. Not to show her that I'm working hard because I wouldn't take her on a trip when there's no choice but to work 12 hours a day. Like the trip to the Carribean would have been ridiculous for her to go on because it was--there were--days began at 7 in the morning and ending at 7:30 in the evening of work. And nothing pleasant at all happening. But if it's going to be for a meeting situation where the amount--actual work is marginal and there are spouse affairs going along with it and there are maybe even some social affairs with it a dance or dinners or things of that sort. And it's apparent that a wife would be--or a spouse would be very much welcome and comfortable there then yes, you know, by all means we've gone into debt over those. I mean we really have. We've spent more than we should have often on those, without a hesitant mind, and I still don't look back on them and think that that was money wasted. By no means was that money wasted.

They have "made a lot of wonderful trips" which have "been one of the highlights of our marriage." She in turn, jumps at the chance to go on a trip with him, and insists that they find a recreational activity to do together--the one they have settled on is ice skating

6H-2: Because at the time we were sort of looking for something to do together and Shirley--the options she had brought up were unacceptable to me, particularly the one to go square dancing. That was one that I had just--we tried some sports, we did volleyball for a while but then I just hurt my shoulder. And I couldn't play volleyball anymore. But we--Shirley said, I remember her saying, "We don't do anything together," or something, you know, she had one of these--and I was, I said, "That's crazy." You know, "We live together we have all sorts of things we do together all the time," but she wanted something specific she could put her finger on that we did together.

Just as forced separations are met by attempts to remain tied together at a distance, unavoidable occasions when pleasurable experiences cannot be shared are countered with efforts to share these experiences vicariously:

6H-6: If I go to a show I--once I've gone, you know, once it's too late, you know, not to go. I try to tell her as much as I can to share with her. She spends a lot of time looking real sad when I'm telling her about it. To make sure that I know that she would prefer to have been there but I do tell her. In--and neat restaurants that I've been to or something like that. Places that I would like to, in the future--us, you know, have us go to together.

Just as he filled his overseas letters with details of his life in the Navy, he shares with her all the details of his business trips--"And the trip is something to be discussed with her as far as I'm concerned"--as well as the ins and outs of office politics. A media specialist for a large company, he brings all of the films he has a hand in producing home for her to view. She always presses him for more details of his solitary trips:

6H-6: My memory is not as good as hers for details. And she always wants more details and usually what happens to me is I don't remember. As soon as I get back I have trouble with some details but over the next week or two weeks or maybe a period--a long period of time as I remember more details I tell her. But I tell her about things like the politics and the meeting itself and what happened and she's really my sounding board on a lot of that stuff. When they've been difficult meetings she really is a sounding board. I've always shared my work with Shirley.

A potential of being joined to another person is that that union can become even closer over time, and Alex characterizes his marriage in this way. He talks of how they are "much more tied to each other now than we were then. We were still two independent people," implying that now they are not. He feels "more in love"; romantic love has "deepened and grown stronger." They like to be with each other more, and also to do things for each other more than before. To separate himself from his wife and children becomes more and more difficult:

6H-4: I just don't think marriage is about me. I think it's about us and each time you add another person in there that person becomes very deeply involved.

. . . . . .

Adding children. It's hard for me to see how you can separate yourself from all of these other individuals that you've become so tied to. Other marriages may not be like mine and Shirley's where I feel, you know, that we are this close.

At some point, and increasingly as his story moves toward the present, he begins to talk about the marriage itself as no longer "two independent people" coupled together, but as a person in its own right:

6H-1: When Shirley came back we shortly after that got an apartment and I think that was the beginning of some good things. We were really more upset about the ship's leaving than we had been about anything else and I think that was a sign that we were beginning to really feel for our marriage.

In a series of similarly striking metaphors, he speaks of the marriage growing and maturing,

6H-6: Our marriage and our love has grown a great deal over the years and only in the last maybe eight, nine, ten years do I feel that it's really matured.

and as something he has a confidence in:

6H-2: I think we pass the crisis then and I had a confidence in our marriage then which I think I was lacking right at the point when we were--you know, well like I said we--I don't think we got to know each other in that very intimate way until we got overseas.

Elsewhere he speaks of some things as being "a breach of the faith of the marriage," more or less "good for a marriage," "helpful to a marriage situation," "hard for the marriage," and "challenging to marriages,"

6H-4: There is certain kinds of analysis, encounter groups and all sorts of things are done now that do focus on the individual. Yes, and I think that is challenging to marriages. What it really comes down to, the big question is, should the couples that break up, as a result of having these things, should they have been married at all? Maybe those marriages weren't any good anyway. But you then begin to wonder--take this a step further--you wonder if any marriage would make it given the--given egotism. If you put the self ahead of the marriage, if any marriages would last. What I'm saying is that I myself don't know whether analysis destroys that type of relationship, or whether it just points out the problems with that type of relationship.

Here the goal of marital permanence is recast in terms of human survival--Alex wonders "if any marriage would make it." And another goal is introduced, that of unselfishness toward one's spouse. This is expressed elsewhere as "the desire to give more to the other person than you're giving to yourself, at times" and "the need to give in the marriage." Egotism, by contrast, is "putting the self ahead of the marriage." The exact logic of unselfishness as a goal of marriage as a PERSON is spelled out in the following statement concerning the effects of his wife's analysis; this statement makes use of the PERSON metaphor once again:

6H-4: I certainly believe that Shirley as an individual should have her rights and privileges as I think I should. But I also believe very strongly in Shirley and Alex as the married couple or the D. family as it exists here as the family that in itself has rights and privileges that must be seen to.

Individual needs may interfere with the needs of the marriage. And analysis, Alex believes, will create this interference because it promotes individual needs:

6H-4: Our marriage is a very good thing for both of us. Has been a good thing for both of us. And is the--to me is the best thing in the world for both of us. And in analysis you have to deal with 100% of yourself, me. You know, the "me" generation is often brought up. And one of the things that's brought up is the--a lot of analysis. A lot of these, "the importance of looking out for Number One" type concepts. Well that's what analysis is to a great degree. You're really going after yourself and your analyst is always saying, "You have a right to this, you have a right to that, you have a right to--" You know, "Other people have no right." Well that really works very well when you're in analysis. It doesn't work very well when you're in the middle of a marriage. Because a marriage is not that. A marriage is something else. It's something where you have to be--have to give away part of this right to always react, certainly.

Therefore Alex is afraid of analysis; for analysis threatens marriage by blocking the marital unselfishness goal in the same way as his wife's flirtatiousness threatened the marriage by blocking the goal of exclusivity, and extended separation threatened the marriage by blocking the proximity goal:

6H-4: When Shirley was in analysis for a short period of time. When that happened that was very hard on me and, you know, we had a very good marriage. And I think analysis--I was afraid analysis was going to ruin it. And I suspect that Shirley is too. Because she said to me, "I dread the day you ever go into analysis because I don't know if you'll--you know, I don't know what's going to happen to you." You know--and--as if something might happen to our marriage.

At the same time Alex agrees that individuals have the right to find out who they are and what they want out of life, and he wonders whether there isn't some alternative way "to analyze people in a marriage as opposed to analyzing people who are not in a marriage situation." Because, as he says about individual analysis, in one final metaphor,

6H-4: If you start to take one part--it's something like trying to analyze only one side of my brain. To take one part of the family and deal that way.

I am arguing that metaphors engender goals. MARRIAGE IS BEING A COUPLE engenders the marital goals of permanence, exclusivity, proximity and shared experience, because these are properties of BEING A COUPLE which realize the chosen metaphor. This metaphor engenders the further goals of connectedness during separation and vicarious sharing of separate experiences because these are alternative ways of realizing the goals of proximity and shared experience, respectively. In the same way, MARRIAGE IS A PERSON entails the goal of marital unselfishness. But marital metaphors, themselves, are organized by underlying conceptualizations. The logic which links these two metaphors seems to be a processual one--that of two independent people coupled together and merging into a single person.

A further set of metaphors ubiquitous in Alex's thinking, marriage as some kind of MANUFACTURED PRODUCT, is linked to the metaphor, MARRIAGE IS BEING A COUPLE. This is done by equating one particular property of BEING A COUPLE--permanence--with a property taken by MANUFACTURED PRODUCTS--durability. The reasoning is revealed in one of the interview segments above:

> 6H-1: That, you know, we were now a family. Just the two of us, we were a family, and that--and the commitment was to make that family move. You know, that it was going to be a solid thing

"just like the parents' had been," with an ensuing discussion of marriage in the '60's contrasted with the mass divorce of today. Another passage which makes explicit the relationship between product manufacture and permanency is

> 6H-1: When we got the apartment in Virginia Beach in Norfolk it was really nice. And it was--I have very fond memories of that place and our marriage at that place. I think we began to really form a marriage there. The fact that Shirley got pregnant also indicates we were getting into something more permanent.

Thus strong or solid or good marriages, or what Alex elsewhere calls successful marriages, like his own, are those which are permanent. His comparisons with other couples inevitably employ the DURABLE PRODUCT metaphor to distinguish between these marriages and others which are "weak," or "broken":

> 6H-2: And we were relying on the kind of looking at each other and saying, "Well," you know, "Who are you?" Now people were doing that up there, married couples were doing that quite often and it resulted quite often in broken marriages. There was a lot of trouble with that. A lot of couples who came up there [TO ICELAND] had problems. And some other couples that came up there seemed to be very strong in their marriages, when they left there, as I think we were. And we remained good friends, although long-distance friends, with some of these couples. That we knew up there who were our age, who had good strong marriages, and whose marriages were obviously strengthened by the stay there. Or at least not weakened by it. I think our marriage was strengthened there. I think we did--we were forced to look at each other very closely.

Strong marriages are less likely to break than weak ones. But the introduction of the DURABLE PRODUCT metaphor seems to permit a proliferation of metaphors expanding this notion of marriage as some kind of MANUFACTURED PRODUCT. A common one Alex uses is a metaphor of marriage as a production or project, the effort at making the marriage, an effort which in his words either "works" or doesn't work, and which you have to "work at" and "work out." In one version, the result is a decidedly home-made product:

> 6H-4: Maybe this is just for us. Maybe this has nothing to do with anybody else. But it could be that--our situation when we got married was such that we had lots of room to adjust. Because we didn't have any idea what we were getting into. That gave us a lot of room to adjust. And by the time we had been through the first year we realized, you know, there would have to be adjustments made. And a few years afterwards when things really got serious we were--you know, when the marriage was strong, it was very strong because it was made as we went along--it was sort of a do-it-yourself project.

While Alex is "not exactly sure why one marriage works and another doesn't," he does have ideas about what has made his own marriage, and other strong marriages he has known, work:

> 6H-2: Well I'd say there's something about those people that we knew, they had a basic solid foundation in their marriages that could be shaped into something good.

Here the product is an edifice of some sort, to allow the notion that what is critical is a solid foundation. The foundation, he believes, is "commitment":

> 6H-2: The couples that we knew that it was working, I don't think there was any question in their minds about getting married. All of them were pretty sure that they wanted to get married. And they were pretty sure they knew what to do with marriage. And this was true about us too, even though, I've said over and over again, you know, how much we've changed and how much our relationship has matured.

(Here he refers to the MARRIAGE IS A PERSON metaphor, which is not applicable in this context),

> I still, at the very beginning, I said, you know, right away we knew there was a commitment. And this was the same thing that was basically true about all these couples. I think they all got married thinking that this was "the thing." The right thing to do and it made sense to them.

In his own case the commitment to marriage is attributable to background, and it becomes clear that the commitment itself is about not giving up the marriage; now, perhaps, permanence is equated with the tenacity of the effort rather than the durability of the product:

> 6H-1: I think we're just lucky as hell, that we have a good marriage. And I think it's not just due to luck but also the fact that both of us are--have deep commitments to the concept of marriage based on our backgrounds and upbringing. We couldn't have given this up once we made the commitment to it, without tremendous trauma.

He attributes this attitude to the common cultural heritage he shares with his wife, because "Jewish families are strong and such" and "the cultural heritage gets carried on." Even though his parents' marriage, unlike his own, is far from ideal,

> 6H-7: I think they believe in home and family too and perhaps I've gotten this strong feeling about home and family from the fact that they're together despite all this bickering and arguing, you know, they would never think of divorcing, I don't believe. I don't think it ever crossed their minds.

But there are other factors which worked for he and his wife:

> 6H-2: I will only say this, what worked for us is, we have a common cultural background. We have the Jewish heritage which is very strong--that heritage is very strong and has made a difference for us. We are supportive of each other and we are *very*, *very* complimentary to each other.

He elaborates on this latter "asset":

6H-4: I think that we were so different and we had such complimentary differences that our weaknesses--that both of our weaknesses were such that the other person could fill in. And that quickly became apparent to us that if we wanted to not deride the other person for their weaknesses we would instead get their strengths in return. And that's what I think has been the asset--these are the assets that have been very good for us. And I suppose what that means is that we have both looked into the other person and found their best parts and used those parts to make the relationship gel. And make the relationship complete.

Here he moves to another PRODUCT metaphor which suggests cannibalizing parts from two broken-down automobiles to build a working one. Elsewhere, the PRODUCT idea allows still another metaphor:

6H-4: I'd certainly hate to see people--everybody in the world jumping into marriage immature. But maybe there's really something to the idea that it is not so much how you feel--exactly how you feel before you get into a marriage but what you can make of life, you know, in the marriage that really counts. Having thought it out--in other words, having thought it out beforehand and coming to the conclusion that you really are in love. Might not be as good for a marriage as having gotten married, looked into what was worthy of being in love about, found it, identified it after awhile, because it does--it takes a while, and then made that the cornerstone.

And finally, another element in their situation which, coupled with commitment, made for a permanent marriage was the very lack of any idea, at the outset, what marriage was about; this understanding permits him still another DURABLE PRODUCT metaphor:

6H-4: But I think that my commitment to the marriage was just so strong and I--as I said before I think this had a lot to do with heritage and background and what we knew had to be done. My commitment to the marriage was so strong that even though, or maybe--I don't know, maybe because I hadn't thought about this, because I had to make it a struggle, or because it was a struggle for us that we forged a lifetime proposition. And maybe it was because things were wrong at first, and we were screwed up and we didn't know what we were doing. Maybe that had something to do with what was good about it. The fact that we really had to work at it.

So that all these different metaphors--of carrying out a project, shaping a product, putting it together from different parts, building an edifice with a solid foundation and a cornerstone, forging something that will last a lifetime, something strong and unbreakable--are permitted by the underlying, and more general, understanding that MARRIAGE IS A MANUFACTURED PRODUCT. And each of these metaphors introduces its own marital goals: you have to start out with the intent not to give up (the foundation); you have to be willing to work and struggle to produce it (the effort); you have to find something worth being in love about (the cornerstone); you have to identify each others' strengths and overlook each others' weaknesses (the parts), and so on. In Schank and Abelson's (ibid.: 116-117) terms, these would seem to be Instrumental Goals--they are all in the service of constructing a durable product. In turn, the conceptualization of marriage as a PRODUCT is linked to a prior understanding of marriage as BEING A COUPLE, via the association between product durability and marital permanence. Unlike other goals of BEING A COUPLE, such as proximity and shared experience, permanence is an Achievement Goal. And it is difficult to achieve, something that must be worked at, a feature which lends itself to characterization in terms of production. Thus, one model, of what marriage should be like, leads to another model or theory, about how it can be implemented. Each suggests apposite metaphors, and each new metaphor is capable of introducing new marital goals.

This highly individualized and relatively complex model of marriage contrasts with the cultural stereotype with which Alex entered marriage--the goal subsumption model which cast marriage as a RESOURCE and asked what he "got out of it." In general the preconceived ideas with which husbands and wives enter marriage, by comparison with the understandings which they ultimately construct through the process of encounter and assimilation and realization, are very sketchy, static, culturally standardized and readily dispelled. These more complex understandings supply any number of simple stories for a story understanding task, such as

Alex decided it was time for Shirley to stop fooling around with any other guys. He got married to her.

or

Alex thought that two years apart would do severe damage to his marriage. He decided to take the post which provided housing for families.

or

Alex feels that analysis focusses on the individual. He was afraid that his wife's analysis would ruin his marriage.

and so on.

The degree to which marital goals are tied to the metaphors which engender them is more obvious by comparison with a different set of metaphors. For this purpose the views of another husband will be somewhat more briefly summarized and interpreted. This husband, Bob, began marriage with a version of the LIVE HAPPILY EVER AFTER myth of American marriage,

5H-2: I don't think many issues did come up. I think that I was a very adept liver in the sense of I did my thing rather well. I was much into being, I think--playing a great deal on the surface of life and living. I think that, it seemed to me that people got married and lived together in a certain amount of harmony, ergo, we will marry and live together in a certain amount of harmony. We didn't pursue in any depth any issue that I recall. That if, I don't know, me walking in with my feet sloppy ticked off Eileen then that was a reason why I didn't walk in with my feet sloppy.

This model was to be jeopardized, first, by their friendship with another couple who were continually questioning their own relationship:

5H-2: I think that we saw at that juncture, probably not even consciously, the potential for a different kind of marriage--a different sort of motif. And I think we were both impressed, awed with it and frightened of it. We didn't want to leap in into an inquiry method of marriage at the point--I don't think either of us felt an impetus to do it, in a sense that it seemed to be the way to a good marriage.

Ultimately, however, a series of extramarital sexual relationships with other people led them to adopt the inquiry method. These relationships called into question, for each spouse at different junctures, the status of their relationship with each other. When Bob reacted to his wife's affair by declaring his own intent to figure out "where he is" on his own,

> 5H-4: She then became very concerned about, "My God. All this means all this. You know what I really care about is my relationship with Bob and I don't want it to go to hell." And I think that was a very important thing to hear for me. And I think from then on, we talked a great deal more about our--where are we, and not where are you but--or where am I vis-a-vis life and love and the pursuit of happiness, but where are you and I?

When he gets involved in an unexpectedly intense love affair, he and Eileen go to the mountains to work through the questions,

> 5H-5: "Where was I? What did it mean?" "What did it mean to the two of us?" Eileen basically supportive and in favor of it but also scared to death and not knowing and I guess in a sense, at least what I--was conveyed to me was, "I kind of--I thought another relationship would be nice and I'm glad that you're having it. But my God I didn't think it would be quite this intense," and all of that sort of thing. The--I think she was not overly threatened by my sexual involvement with Martha but that was a weighty dimension. The--I think she was more concerned about the depth of our love and the basicness of my love at that point. Which again, excluded Eileen but--certainly had an intensity that was different from my love for Eileen at that time.

It is not only Eileen who is scared by the depth of Bob's involvement with Martha:

> 5H-6: I think the risk is basically that the situation of the relationship with Eileen and I would get out of hand, such that I couldn't insure that it would continue no matter what I did. And so I would be in a position of losing someone that I thought so much of and I could do little about it. And so if I played it safe, if I played it tentatively, I had a certain control in it, in the sense that Eileen could be running around and making a risk of herself in a sense, but if I didn't I could always hold it together and keep it going and all that. Which I might think retrospectively is not probably right either but that was my perception. So I think really what was at stake was the--was ending--losing the relationship.

Bob and Eileen meet each one of these marital crises by talking:

> 5H-4: I think it worked out that Eileen and I both perceived that we had little idea of, in fact, what our commitments to one another involved. That we did not in many ways know ourselves particularly well, relative to relationships, relative to life and living and that sort of thing. I think it worked out that through miles of talking, with one another, on a lot of this, that it at least made sense to us and felt important to us to continue our commitment and our living together with one another.

And on another occasion,

> 5H-5: So we went back to the mountains and I think it was a time of a lot of walks. It was a time of a lot of searching. A time when we gave each other some space to kind of work out where each of us were....I think at the end of that time together, we both were--it was pretty clear to both of us that our love was a very strong and deep one.... And in that it was a biggie, that we really had to keep in contact about it, we really had had to talk about it.

The conclusion which emerges "through miles of talking" is that the relationship with each other is a basic or PRIMARY RELATIONSHIP, a "biggie," and this understanding makes sense of their relationships with other lovers:

> 5H-5: There was a primary commitment to me and the relationship with Dave was ancillary, may deal on areas that I didn't. But was largely additive and so that made it good and okay and reasonable and all that. That's kind of a rule or an understanding in much of that that made sense.

> 5H-9: And so there's been a very strong draw. I think that as we were slugging through some of the more rational issues or those that we could get into some sort of rational component, what really prompted us to do that work and what really made it important that we come to some clarity is the strength of the emotional sort of thing. Which again I think we both recognized and then just a--I don't think we glorified it as much as acknowledged that it was there. That is to say my hurt hurt Eileen, her hurt hurt me and I can be hurt by lots of other people's hurt but I think not to a depth that that seemed to consistently have, and then with her as well. There--maybe it's the combination that there is a--there's an intellectual stimulation with one another, there's an emotional stimulation with one another, there's a sexual stimulation with one another, there's a childbearing stimulation with one another, or wrestling with great issues of the world, and so I think Eileen encapsulates for me an ongoing growth potential for me and all that gambit and vice versa, I believe. So--and I think we have found parts of that in many other people many times, but no one who we felt could replace in that sense.

MARRIAGE IS A PRIMARY RELATIONSHIP, comparable to other relationships, but more basic, important, emotionally deep, and irreplaceable. Unlike Alex's model of marriage, it does not entail exclusivity. Consistent with this view of marriage, Bob reports that he has trouble with the term 'marriage' itself, and he is far more inclined to speak of their 'relationship.' He comments that

> 5H-8: I think that we looked at our relationship very often in contexts of other relationships. We have been as a couple at times very involved with other friends and what they were doing or not doing. And they have seemed very comfortable sharing with us where they are and where they aren't and I think we have provided for them sometimes new vocabulary and vice versa. I think Eileen has been more open about where she is with her friends than I have. Getting out types of new vocabulary or why don't you look at it this way. Or getting support for areas that she was in some sense holding out or trying to say to me, "No, it needs to go another way." That sort of thing. So I think other people provided us both through their direct request for *our* counsel and help and all and the times in us utilizing their marriages as kind of case studies for where we are in all this.

This use of other marriages as "case studies" providing new vocabulary and new perspectives helpful in the task of understanding their own relationship, contrasts with the way Alex uses other couples' marriages. For Alex, other marriages are either like or unlike his own in terms of durability, and their comparison provides him with some understanding of what that difference is.

If MARRIAGE IS A PRIMARY RELATIONSHIP, the Preservation Goal is to keep this relationship together in the course of other, less important ones. But a RELATIONSHIP does not link two people together physically, as does BEING A COUPLE; rather, it positions them vis-a-vis one another. Thus the problem of implementing this goal is not one of constructing a more durable product; it is one of "keeping in contact," a major Instrumental Goal of this marriage. This involves the "inquiry method" of marriage. Marriage as INQUIRY lends itself to the further metaphor of mutual self-examination, and hence marriage, as a SPATIAL RELATIONSHIP.

> 5H-7: I don't know I think that if she were to change dramatically or I were that that may not be so. Perhaps part of our commitment to continually evaluate where we are in some sort of awareness of that phenomenon. That the greatest liability to our relationship is to not work on trying to get a good sense of where we are. If I let her alone and don't try and she lets me alone we might satellite far enough away so we're not sure what's in between us.

This last is a kind of "Lost in Space" metaphor which nicely captures the sense that there are no reference points except each other. In this spatial metaphor, "keeping in contact about it" means "talking about it." Talk, which initially served to establish the primacy of their relationship to each other, now serves to keep each spouse apprised of where the other is so that they can, presumably, do whatever necessary to "hold together" their relationship.

> 5H-5: So I think that the relationship with Martha probably set a lot of ground rules for how we're going to communicate through other relationships, at other times. I guess it also very deeply confirmed that we can deal with and process in the throes of living relatively effectively, a lot of stuff. And that if we don't, if we try to pack it off somewhere, very quickly there's a very perceptiv--perceivable sterility, by each of us, at where our relationship is.

Again, the role of communication in this marriage contrasts sharply with the use of talk by Alex and Shirley to convert unshared into vicariously shared experience by telling all the remembered details of their times apart.

Alex feels his marriage threatened when certain conditions--his wife's flirtations, her analysis, their lengthy separation--arise to block critical marital goals. He generally responds to such situations by doing his best to eliminate the threatening condition. What threatens Bob's marriage is a much more gradual process of change in one or the other spouse, which can be brought about by new relationships or new experiences of any kind, and which is so nearly imperceptible that it requires continual monitoring:

> 5H-7: I guess what I'm saying is I don't know how big a change that I thought there might have been but I think that there is always a thought when Eileen goes into a different sort of experience and--or I do, of the other partner saying, "How much will this change them?" Indeed, Eileen's going to her job in Greensboro. There was a part of me that monitored very carefully how she was changing in the aspect of, "Is this still the person who I want to interrelate with? Can I deal with this change? Does it seem reasonable to me? What do I have to gain or lose from it?" That whole bit.

Such separate experiences cannot be dealt with, simply, by sharing them vicariously, for they alter the individual in ways that may simply move him or her too far away, so that the relationship can no longer be held together. Efforts to keep in contact may not be enough, either. Changes in the other person may be so great as to call into question whether this is "still the person who I want to interrelate with"--whether this should continue to be the PRIMARY RELATIONSHIP. For unlike Alex's version of a marriage as some kind of physical coupling, a PRIMARY RELATIONSHIP cannot necessarily be expected to be permanent.

> 5H-7: I think that what our relationship is, is important because we're two other people that *do* live together, we *have* lived together, Eileen and--the information and the knowledge and the respect and love for me that she has is enormous and vice versa. And I think that now it's much more a commitment to making certain that we're pretty clear where each of us are in our growth, in our lives, in our living. Not, I guess--what I don't feel in what you asked is, not to keep the relationship together but because we have a relationship. I *truly* could conceive of us in our evolution of things feeling that an ongoing relationship, living together, does not make sense. When we're at a point in growth and who we are or because of another person or another opportunity that seems to be there that says, "Okay we need not and we probably should not perpetuate this." I guess here again I feel very simple-minded about the whole thing in that that we live together we need to clarify these things there's a compelling interest to do so. I feel pretty much the same with most everyone as a matter of fact, but not to the depth of commitment to doing it. I mean it's very important to me that Eileen and I are pretty clear where each *of us are.*

So that the final Instrumental Goal of marital self-examination is to assess when it is time to discontinue the relationship, either because the two spouses have moved too far apart, or alternatively, because another relationship has taken primacy. Such assessment, these passages hint, is understood within yet another metaphor, of gain, loss and substitutability--A RELATIONSHIP IS AN ECONOMIC EXCHANGE.

Elsewhere (Quinn n.d.) I have described the role of the word 'commitment' in framing marital goals. In one of its senses, marital commitment is to those goals which are effortful and ongoing enough to require substantial dedication to them. It is not surprising, then, to hear Alex say, "....the commitment was to make that family move. You know, that it was going to be a solid thing." while Bob speaks of "our commitment to continually evaluate where we are," "a commitment to making certain that we're pretty clear where each of us are in our growth, in our lives, in our living." The word 'commitment' is used to mark those marital goals which are instrumental to the implementation of a given model of marriage.

I hope to have hinted that two-liners from story-understanding tasks are poor sources of inspiration for a theory of goal organization. All of us can interpret any one of the Alex stories I invented above, and probably for the most part interpret it correctly. This is because we have a great deal of knowledge about the possible links between particular goals and

particular actions, including both knowledge of human nature and of American culture. This knowledge helps us understand what Alex might be up to in the particular instance; it does not tell us much more, because to do such spot interpretation of someone else's behavior we do not ordinarily need to know his philosophy of marriage. We do, indeed, have some folklore about the way in which marriages may organize goals. This folklore again, is a misleading source of information about the way goals are actually organized in ongoing marriages. It may tell us a lot more about why John decided to get married than about why John decided to get divorced. The organization of marital goals and the goals of other interpersonal relationships is better understood in the richer context of extended discourse about the discourser's own relationship--that is, from the inside. While such discourse is not perfectly transparent by any means, it does suggest a view of goals as organized by metaphors which are themselves organized around underlying theories of what a relationship should be and how such a relationship can be implemented.

REFERENCES CITED:

Quinn, Naomi
n.d. Analysis of a key word: 'commitment' in American marriage. Unpublished ms.

Schank, Roger and Robert Abelson
1977 Scripts, plans, goals and understanding: an inquiry into human knowledge structures. New Jersey: Erlbaum.

Wilensky, Robert
1978 Understanding goal-based stories. Doctoral dissertation, Department of Computer Science, Yale University.

SYMPOSIUM
COGNITION AND PERCEPTION

TRANSFORMATIONAL STRUCTURE AND PERCEPTUAL ORGANIZATION

Stephen E. Palmer
University of California, Berkeley

The problems of structure and organization in perception are among the most central in the history of the field. Beginning with the gestalt movement early in this century, researchers and theorists alike have supposed that certain types of perceptual phenomena provide insights into the structure and organization of the system that produces them. There are two major questions. First, what is the common element in these phenomena that provides the most transparent view of the internal structure of the system? Second, how is this structure embodied and used within the system so that it produces the observed phenomena?

In this paper I suggest answers to both questions. To provide an initial idea of the direction I will take, here is a brief preview. The organizational phenomena I take to be most critical for understanding the structure of perception are: (a) object constancy despite changes in image size, shape, position, and orientation, (b) motion perception of structured objects, (c) figural goodness or "good gestalt", (d) perceptual grouping of the visual field, and (e) reference frame effects. First, I will argue that these phenomena are related by the fact that underlying them all is a common transformational structure. Second, I will describe a general design for a computational system within which transformational structure can be analyzed easily. The proposal is to couple a transformationally based system of analyzers with an attentional mechanism that establishes a variable frame of reference within it. The job of the attentional frame is to move around within the space of analyzers so as to maximize the invariance and stability of the attended output. It does so by compensating for the changes brought about by the transformations. Finally, I will make a few remarks about why I think the kind of computational structure I am proposing is an interesting one from a purely theoretical standpoint.

Transformational Structure

The basis of transformational structure is the concept of transformational invariance. Transformational invariance refers to the fact that when an object undergoes a spatial transformation, such as a rotation, a great many changes occur in the pattern of stimulation on the retina, but at the same time there is a great deal of higher order structure that does not change at all. All sorts of relationships (or relationships among relationships) do not change, even though each first order property does, and these unchanging aspects are the transformational invariants. In many cases they correspond more directly to the intrinsic properties of the real world object undergoing the transformation -- its size, shape, color, and so forth -- than to the properties of its projected image. The image, after all, changes in ways that the object clearly does not. Because transformational invariants reflect object properties rather than image properties, computing them probably plays an important role in getting from an image based representation to an object or world based one.

How can the transformational structure of an event be computed from the spatio-temporal images that arise from it. The problem, of course, is that while it is trivial to compute the image over time from knowledge of the real world object and its real world transformation, the reverse is not at all trivial. In fact, there isn't even a unique solution, but only an infinite set of object/transformation pairs. In other words, there are many different world events -- objects undergoing transformations -- that could give rise to the same specific event images and no logical basis on which to decide the correct one from purely optical information. It is clear that some additional assumptions are needed about either the nature of the object, the nature of the transformation, or both to reach a determinate solution. To reach the <u>correct</u> solution as well, the additional assumptions will have to be chosen in a principled way.

What assumptions might help here? Clearly a good bet would be any assumption that is generally true of events in the world. Then, arriving at solutions that are consistent with these assumptions will almost always result in veridical perception.

Perhaps the most striking fact about the transformations that characterize events in the world is that both objects and transformations tend to be fairly stable over time and space. This is certainly true over short intervals of time and small regions of space and is usually true over long intervals of time (on the order of at least whole seconds) and larger regions of space as well. To the extent that this is so, objects can be well approximated as rigid and the transformations they undergo as uniform motions in three dimensional space. Naturally, there are many cases of non-rigid objects undergoing various complex motions. But even these are probably best understood in terms of how they can be analyzed into a structure of roughly rigid components undergoing roughly uniform motions. The motion of a person walking is a case of non-rigid motion that has yielded to such an analysis (Johansson, 1973; Cutting, 1981).

The rigidity and uniformity assumptions suggest that the perceptual system operates in such a way as to <u>maximize</u> invariance in both objects and motions. In anthropomorphic terms, the system "wants" to analyze both objects and motions as "changing as little as possible". "Wanting to change as little as

possible" refers simultaneously to the facts that that the perceived object/motion pair must, in some sense, account for the sensed variations in the image and that this can be done in more than one way. To pick a well studied example, the kinetic depth effect, if a line changes simultaneously in its projected length and orientation, it might be either (a) a line that gets shorter and longer by certain amounts as it rotates in varous possible ways or (b) a line of constant length that is rotating in depth (Wallach & O'Connell, 1953). The latter is almost invariably perceived, although it sometimes takes an observer several seconds to achieve it. Once it is perceived, however, it is stable and does not spontaneously change to something else. According to the present line of thought, the preferred interpretation arises because it is "simpler" than the alternatives given the heuristic assumption that objects in the world tend to be rigid and undergo uniform motions in three dimensional space. Thus, the perceptual system seems to prefer interpretations of greatest possible invariance.

Motion and Constancy

The perceptual phenomena most directly and obviously related to transformational invariance are those of motion perception and object constancy. Motion occurs when the position of some object or part of an object changes over time. It is the paradigmatic case of a transformation, of course, but it is perhaps not so obvious what it has to do with invariance. In fact, the whole concept of a distinct object undergoing motion presupposes invariance in that the object is taken to be unchanging (except for its position, of course) as it moves. Logically, one could just as well say that the world had changed its intrinsic nature over time. This would be a more reasonable notion if one perceived the visible surfaces of the world like a rubber sheet that simply changed its shape plastically during events. The fact is that people to not perceive the world in this way, but as consisting of articulated, unchanging objects that undergo various sorts of motions. This highlights the fact that perceiving motions of objects actually presupposes invariant aspects as well as varying ones, and that an event in the world always has both components.

It appears, then, that object constancy is just the other side of the coin from motion perception. Motion is the perceived transformation; object constancy is the perceived invariance. They are completely coupled in that for the motion to be different, the object must be different too. In the case of a rotation in depth, for example, either the object is rigid and the motion is uniform (as in actual depth rotation) or the object is plastic and the motion is nonuniform in such a way that their combined changes produce the two dimensional image (as in the plastically deforming colored regions in a motion picture of a depth rotation).

Given that real world events tend to consist of rigid objects in uniform motions, a perceptual system would be biased toward veridicality if it somehow embodied preferences toward perceiving rigid objects and uniform motions. Indeed, there is good evidence that this is true. When presented with ambiguous information, people tend to perceive rigid objects rotating or translating in three-dimensional space as long as the optical structure is consistent with such an interpretation and the system is given sufficient time. For example, Johansson (1950) showed that people have a strong tendency to see two moving points as fixed at the ends of a rigid rod moving in three dimensions rather than as moving non-rigidly in two dimensions. Thus, an ambiguous object in unambiguous motion tends to be perceived as rigid rather than plastically deforming. The other side of this story is that an unambiguously specified object in ambiguous stroboscopic motion tends to be seen in uniform motion (Shepard & Judd, 1976; Farrell & Shepard, 1981). Clearly, there is something special about rigid objects and motions for the human visual system.

Figural Goodness

Another important problem in perceptual theory that is intimately related to transformational structure is what psychologists have come to call "figural goodness." Figural goodness refers primarily to subjective feelings of order, regularity, and simplicity in certain figures as opposed to others. The relation of this subjective feeling to transformational structure is not intuitively obvious, but it is nevertheless simple to grasp.

Figures are "good" to the extent that they are themselves invariant over certain types of transformations. The most obvious case is that of standard bilateral or reflectional symmetry. To illustrate, the letters "A" and "M" are reflectionally symmetric about their vertical axes because each letter is the same as itself after being reflected about a vertical line through its center. The other widely known type of symmetry is rotational. The letters "N" and "Z" are rotationally symmetric through an angle of 180-degrees because each letter is the same as itself after being rotated by 180-degrees. Still other letters have transformational invariance over a number of different transformations: "X", "O", "H", and "I" all have two reflectional symmetries (about vertical and horizontal lines through their centers) as well as 180-degree rotational symmetry. A perfect circle has still greater transformational invariance because it is unchanged by all central rotations and reflections.

There are two other, less well known types of symmetry: translational and dilational (Weyl, 1952). They are defined by the same abstract scheme as for reflectional and rotational symmetries. Both of these latter sorts of symmetries technically apply only to idealized, infinite patterns, but one can define "local" versions that apply to

finite patterns by only requiring invariance for part of the pattern over the transformation. (See Palmer, in press, for a more complete discussion of symmetry, local symmetry, and their relation to transformational structure.)

It turns out that the goodness of figures can be well predicted from its symmetries in this extended sense: the set of transformations over which the figure is invariant. Garner (1974) showed that ratings of perceived goodness increased monotonically with the number of transformations over which the figure is invariant. Garner actually talks about "rotation and reflection (or R & R) subsets" of a figure, but this concept turns out to be isomorphic to the number of rotational and reflectional invariants (Palmer, in press). Further, the amount of transformational invariance a figure has also strongly affects how quickly people can match two figures for physical identity, how well they remember a figure, and how easily they can describe it. Such results demonstrate the reality of figural goodness in perceptual processing. (See Garner, 1974, for a review). There is additional evidence that figural goodness depends on translational and dilational invariances as well (Leeuwenberg, 1971).

In summary, figural goodness seems to be characterized quite nicely by the concept of transformational invariance. The relevant transformations in this case are reflections, rotations, translations, and dilations. Except for the addition of reflections, these are the same set that characterized motion and object constancy phenomena. It seems unlikely that this is merely a coincidence.

Grouping

The next phenomenon I want to relate to transformational structure is grouping (Wertheimer, 1923). Figure 1 shows some standard examples of grouping phenomena in which most people report perceiving either a vertical or horizontal organization. Figure 1A demonstrates the influence of proximity on grouping. The dots are organized into vertical columns (rather than horizontal or diagonal rows) because their vertical proximity is greater than their horizontal proximity. Figure 1B demonstrates the influence of similarity of orientation. All else being equal, similar elements tend to be grouped together and dissimilar elements grouped apart. Many different kinds of similarity have been shown to affect grouping, but color, size, and orientation are particularly striking. Continuity, symmetry, and closure are three other well documented factors. But perhaps the most potent of all is what gestaltists called "common fate." Elements group together by common fate when they move in the same direction at the same rate. Even a completely homogeneous field of random dot texture spontaneously organizes into figure and ground when a spatial subset of the dots begins to move together or when the rest of the dots begin to move around them.

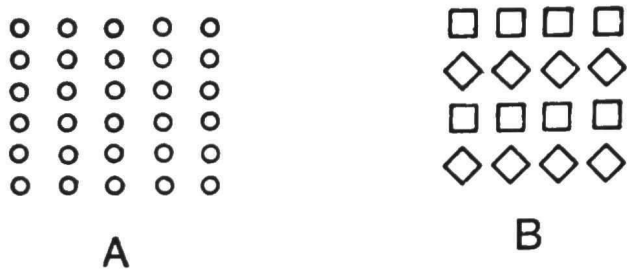


Fig. 1: Grouping by proximity and similarity.

The point I want to make about grouping phenomena is just this: Elements are grouped together when they are in closer transformational relationships to each other than they are to other elements. Transformational "closeness" refers to the magnitude of the transformation required to achieve transformational invariance. For example, the dots in Figure 1A must undergo a larger translation to bring them into congruence with their vertical neighbors than with their horizontal neighbors. In Figure 1B, the figures must undergo a rotation as well as a translation to coincide with their vertical neighbors, whereas just a translation will suffice for the horozontal neighbors.

Now consider some of the most potent factors in grouping phenomena. Similarity of two elements in position, orientation, and size can be defined by the magnitudes of the transformations -- translations, rotations, and dilations, respectively -- that are required to make them equivalent. Continuity is similarity over local translations, and bilateral symmetry is similarity over reflections. And so, once again, we find the same types of transformations lurking behind grouping phenomena as we found behind motion, object constancy, and figural goodness. Grouping seems to be determined by maximizing transformational relatedness within a perceptual group.

Frames of Reference

The final category of perceptual phenomena I want to discusss I will call "reference frame effects." It includes a number of different results in many different domains. What they all have in common is to suggest that perception at any moment occurs within a single, unitary frame of reference that captures common properties of the whole display. Other properties are perceived relative to this frame, very likely in terms of deviations from it.

As an example, consider how the orientation of a global reference frame can affect shape perception. Figures 2A and 2B show the same form in two orientations that differ by a 45-degree rotation. Figure 2A is perceived as a square because its sides are horizontal and vertical, and Figure 2B is perceived as a diamond because its sides are diagonal. However, an interesting thing happens when a number of such forms are aligned diagonally. The perceived shapes

reverse: horizontal and vertical sides produce the appearance of diamonds while diagonal sides produce the appearance of squares (Attneave, 1968; Palmer, in preparation). It seems that the tilt of the whole configuration is somehow "factored out" of the display, and the orientation of the sides is then perceived relative to the whole configuration. The conjecture is that this "factoring out" is done by establishing a tilted frame for the figure within which 45 degrees is the referent orientation. This would explain why the shapes are perceived as they are, and it fits well with many other orientational phenomena in shape perception. (See Rock, 1973, for a review).

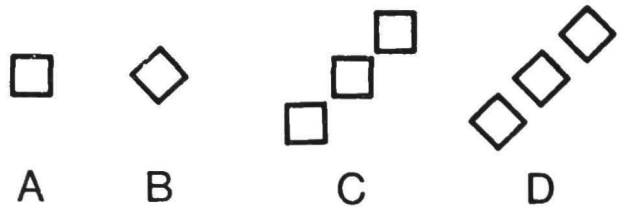


Fig. 2: Reference frames in shape perception.

Similar effects are well known to occur in motion perception. For instance, when two points are in sinusoidal motion, one vertically and the other horizontally as shown in Figure 3A, people do not usually perceive them as such. Rather, the see a configuration that moves diagonally as a unit, within which the two dots move toward and away from each other as depicted in Figure 3B (Johansson, 1950). Here again, it seems that the perceptual system establishes a frame of reference for the common motion and "factors it out." "Induced" motion effects are similar in that an unmoving object is seen to move because a larger, more prominent optical structure serves as the frame of reference, but is moving so slowly that its motion is not detected (Dunker, 1929).

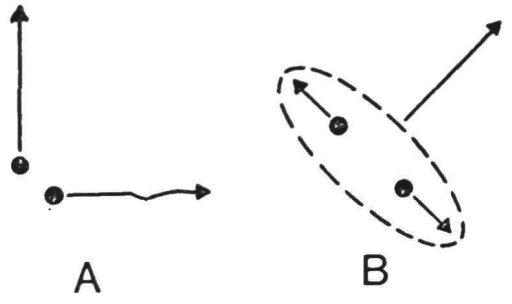


Fig. 3: Reference frames in motion perception

The relation of reference frames to transformational structure can best be explained by analogy to their role in analytic geometry. There, a reference frame is established to describe a point's location numerically. It provides a set of assumptions within which spatial positions maps into numerical coordinates. Different reference frames map the same point into different coordinates, but they are related by exactly the same transformations that relate their corresponding frames. In standard Euclidean geometry, this set consists of translations, rotations, dilations, and reflections. Thus, the structure underlying all possible Euclidean frames of reference is based on transformational invariance: the set of transformations that make one reference frame equivalent to all others.

Returning to perception, it seems that the perceptual orientation of a line or perceived direction of motion is like the position of a point in analytic geometry: their values depend on the reference frames within which they are perceived. Like its geometrical counterpart, this frame seems to include something akin to a position (origin), orientation (axis), resolution (unit size), and reflection (sense). And if this frame is variable from one moment to the next, then an underlying structure of transformations is implied that relate one frame to another. These transformations are exactly the same as we have encountered repeatedly: translations (for position), rotations (for orientation), dilations (for unit size), and reflections (for sense).

Transformational Theory of Perceptual Structure

All of the perceptual phenomena just discussed suggest that the perceptual system has a definite preference for processing optical structure involving certain kinds of transformational invariances. The questions I now want to address are (a) what this might tell us about perceptual organization and (b) how such a system might be constructed computationally.

I think the examples considered above are telling us that the visual system is built on a transformational base that can be used to extract transformations as a heuristic for solving perceptual problems. In other words, the system is designed to be transparent to transformations of the sort most often encountered so that it "prefers" interpretations involving them. By "transparent" I mean that (a) these transformations can be computed rapidly and easily in such a way that (b) the system can compensate for them simply and efficiently. This strongly suggests that the system must be designed to solve the problems of transformational invariance right from the start, and that these design features form the heart of the system.

There are three basic components in the solution I will consider here: (a) a space of analyzers that are transformationally related to one another, (b) higher order analyzers that compute output similarity of lower order analyzers over local transformational relations, and (c) an attentional mechanism that establishes a perceptual reference frame within the analyzer space that maximizes invariances. The output of attentional fixations is stored in memory as a representation of the perceived object. I will discuss each part in turn more fully, but the reader should remember that they are interrelated proposals that only make sense within the complete systemic structure.

First-order Analyzer Space

The first component consists of a set (or sets) of analyzers computing spatial properties of the visual field in parallel. Surprisingly, almost any sort of analyzers will do, as long as they have particular structural relationships to each other. The structural constraint is that they be transformationally related to each other. Precisely what this means is developed more fully and formally elsewhere (Palmer, in press), but the basic notion is just this: Two analyzers are transformationally related if their "receptive fields" or "spatial functions" are identical except for a transformation from a specific set. In the present case, the set consists, not surprisingly, of the transformations discussed earlier: translations, rotations, reflections, and dilations (the so-called "similarity" transformations of Euclidean geometry). I call such sets of transformationally related analyzers functional systems because they compute, in this transformational sense, the same spatial function (Palmer, in press).

An example would be a set of excitatory "bar detectors" with inhibitory surrounds (ala Hubel & Wiesel, 1962) whose elements differ only in the position, orientation, and size of their receptive fields. Each bar detector is related to each other one by a translation, rotation, dilation, or some composite of two or more of these transformations. A similarly constructed set of edge detectors would constitute another functional system, distinct from the first because there is no transformation from the specified set that makes a bar-like receptive field into an edge-like receptive field and vice versa.

The overall structure of a functional system can be conceptualized as an analyzer space in which each analyzer is a point. The dimensions of the space correspond to the transformational relations among them -- i.e., the position, orientation, resolution (or size), and reflection (or sense) of the analyzers relative to each other. Since the number of analyzers is certainly finite, the space is only sparsely populated with analyzer points. Therefore, it is more appropriate to think of it as something like a discrete lattice structure such as depicted in Figure 4. The cyclic dimension is orientation (which repeats after 180-degrees of rotation) and the binary dimension is reflection (which repeats after each reflection), while both positional and resolutional dimensions are simple orderings. The diagram is naturally a simplification of the actual space, since one cannot depict a structure of more than three dimensions in real space. To think of the whole structure in concrete terms, one can conceive of the vertical dimension of the structure shown in Figure 4 as resolution and then imagine a whole two dimensional array of them (to represent the two positional dimensions) like a case full of beer cans. Notice that the relations between pairs of analyzers in the space reflect the transformational relations between them as discussed above. This transformational structure defines the system.

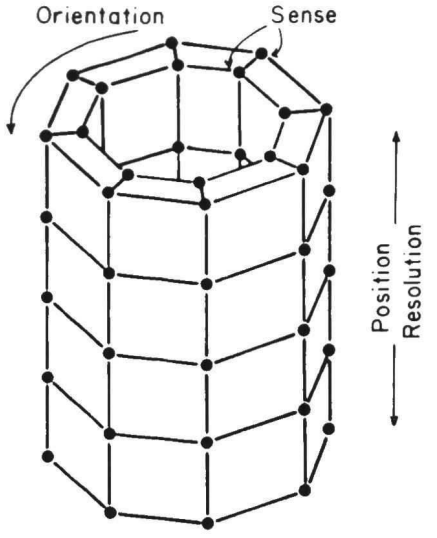


Fig. 4: The space of analyzers.

Higher-order Analyzers

The importance of transformational relatedness among analyzers is that their output is guaranteed to be the same given any two patterns (or portions of patterns) that are identical over the transformation that relates them. For instance, consider a pattern with reflectional symmetry about a vertical axis such as the letter "A". As discussed before, this means that it is invariant over a reflection about this vertical line. Now consider any analyzer that covers any portion of this pattern. Its output, whatever that might be, must be identical to that of the other analyzer related to it by reflection in the same vertical line. Moreover, this is generally true for all pairs of analyzers related to each other by this particular transformation given any pattern having that type of symmetry. Thus, the interesting fact about transformational relatedness among first order analyzers is that transformational regularities in the stimulus event will be reflected in easily computable regularities in their outputs.

In general, higher order analyzers are elements that compute such relationships among the outputs of lower order analyzers. They respond to symmetries or motions, depending on whether they compare outputs simultaneously or over a time lag. We have just discussed a case involving symmetry, but it turns out that motion analysis has exactly the same logical structure except for the introduction of a temporal difference. For example, suppose that a pattern at some time produces outputs in the first order analyzers and that the pattern then moves, say, by a translation. The output this produces after a short duration will be exactly the same as

the output it produced initially for each pair of analyzers related by that translation over that time lag.

These second order analyzers can be conceptualized as the links in the lattice structure depicted in Figure 4, although there probably would be far more of them. They represent local transformational relations among first order analyzers. They compute regularities in space (symmetry) or space-time (motion) by determining transformational invariances. One can even think of them as being embedded in the same space as the first order analyzers because they have the same sort of transformational structure. In the case of motion analyzers, of course, there is the additonal dimension of rate of motion.

This transformational structure of the second order analyzers allows the possibility of piggy-backing still higher order analyzers on top. That is, the outputs of second order analyzers could be compared for similarity in just the same way that they compare the outputs of first order analyzers. This makes most sense for motion analyzers. Those analyzers that compare outputs simultaneously would provide information about symmetries and regularities of motion. Those that compare over a further time lag would provide information about accelerations and decelerations.

Visual Attention and the Mind's Eye

Given such a space of analyzers, how can it be used, as intended, to "factor out" transformational structure? The real problem here is to find some internal transformation that will compensate for the external transformation. One transformation "compensates" for another if applying the second after the first yields the identity transformation -- i.e., invariance or no change at all. For example, suppose an object is coming directly toward an observer. Over time, the image of this object expands uniformly in the visual field. If the observer were to move away from the object at the same rate as the object moved toward the observer, then the two transformations will exactly cancel, and the image of the object will not change. Thus, the "moving backward" transformation by the observer exactly compensates for the object motion.

I want to suggest that something similar happens inside the head. Rather than the eye compensating for transformations by moving about in the world, however, I suggest that visual attention moves about within the analyzer space, playing the role of the mind's eye. Like the eye with respect to the world, visual attention can change its position and orientation with respect to the analyzer space. Unlike the eye, it accomplishes both of these transformations by simple movements within the analyzer space. That is, rotations of the mind's eye correspond to translations of visual attention along the orientational dimension of the analyzer space. Similarly, changes of scale ("zooming" ala Kosslyn, 1981) can be accomplished by translations along the resolution dimension.

Perceptual Reference Frames

At the heart of this proposal is the hypothesis that visual attention is localized within the analyzer space and is centered at a particular position. This establishes a perceptual reference frame for further perceptual analysis. Fixing visual attention on one position of the analyzer space induces a reference frame because it determines a position (or origin), orientation (or axis), direction along that orientation (or sense), and resolution (or unit of distance) relative to which the contents of visual attention are coded. Thus, positioning visual attention determines the description given to objects under analysis. This is entirely analogous to the role of reference frames in analytic geometry. A circle has a different equation when the origin of its reference frame is at its center than when the origin is off-center. The corresponding phenomenon in perception occurs when the same figure looks like a square or a diamond, depending on the orientation of the reference frame within which it is perceived (see Figure 2).

If visual attention acts as a perceptual reference frame for constructing desriptions of shapes, then it is clear that certain rigid transformations can be completely compensated for by corresponding attentional changes within the analyzer space. The result is analogous to a geometrical reference frame displacing as the circle did or changing its scale size as the circle grew larger so that its equation did not change despite the changes in the geometric figure. It is easy to see that the changes in reference frame that would accompany displacements of such an attentional mechanism would be able to compensate for rigid transformations of objects, keeping the contents of the attentional frame -- whatever they might be -- constant despite the transformation. It is not difficult to imagine that such a system would "prefer" to register uniform transformations of rigid objects rather than complex motions of deforming objects. Compensating for motions of rigid objects can be accomplished simply by an attentional transformation and it does not require any change in the description of the object. Any other sort of transformation requires changes in the object's description as well. If the attentional frame somehow manages to follow the path within the analyzers space that maximizes object constancy and motional uniformity, then its operation will embody such "preferences."

Attentional Control

This leads directly to the next problem: how this attentional reference frame is controlled. It should be mentioned at the outset that there is a certain amount of conscious control over the attentional reference frame. People usually can, if pressed, attend to specified positions, sizes, and orientations. But I suspect that conscious control of attention is a high level cognitive activity that does not usually extend down to the level at which we are currently dealing. Rather, it seems that a great deal of the nitty-gritty details of

attentional control must be strongly determined by stimulus structure.

How might this be done? The answer I want to explore is that the structure of stimuli determines how visual attention is positioned -- their symmetries and regularities. The basic idea is that attention is positioned to maximize transformational invariance. This can be done by finding the maximal output from the higher order analyzers for a given region of the analyzer space, since these analyzers are the ones sensitive to transformational invariance.

It is important to realize here that different reference frames imply different symmetries and regularities. Therefore, any "economy of coding" scheme for representation (e.g., Attneave, 1954) requies that the reference frame be chosen to maximize such symmetries. To illustrate this fact, consider again the case of a circle. When the origin of the reference frame is at its center, the circle has all possible reflectional and rotational symmetries centered about that point. When the origin is off-center, its only symmetry about that point is a single reflection about the line joining it to the circle's center. Such facts will be represented in the outputs of the second order analyzers at the center and off-center positions within the analyzer space. The output will be much greater at the position within the analyzer space that corresponds to the center. Therefore, there will be a strong tendency to establish the attentional reference frame at the center of the circle.

For a circle, there will be no particular orientation preference, precisely because it has complete central symmetry. For a square, however, or for any other figure that has significant asymmetries, there will be decided preferences in orientation. A square is symmetric about only the lines joining opposite midpoints of its sides or opposite vertices of its angles. Therefore, only these four orientations are serious candidates. If the midpoint line is used, the figure would have a different perceived shape than if the vertex line were used. In fact, these two frame orientations result in the "square" and "diamond" interpretations, respectively. Obviously, not all figures have exact symmetries like circles and squares do. But the same principles would apply to approximate symmetries.

The notion I have in mind for the placement of attention is a "hill climbing" process. Its goal is to maximize transformational invariance in the stimulus information by seeking the position of highest output in the analyzer space, at least locally. Sometimes there will be several maxima, and in these cases quite different percepts will arise when different maximal positions are chosen for the perceptual reference frame. The square/diamond and ambiguous triangles are two well known examples (Attneave, 1968; Palmer, 1980; Palmer & Bucher, 1981).

A single attentional fixation will seldom be sufficient to code a whole scene or complex object. More complete coding would be accomplished by making many attentional fixations at other positions within the analyzer space that have high ouputs. There is evidence of a bias toward beginning at the most global level (Navon, 1977). A reasonable guess would be that after one or two global fixations of an object at a low resolution level of the analyzer space, many local fixations would be made at higher resolution to code details. I am assuming that the contents of these attentional fixations are somehow stored in memory to form a representation of the world. The result will be a hierarchical structural description, much like those I have discussed previously (Palmer, 1975, 1977).

Given that attention can be positioned to maximize transformational invariance, it is not difficult for it to continue to do so if the object begins to move. That is merely a matter of tracking the same maximum through the analyzer space with the help of the motion analyzers discussed earlier. Recall that these analyzers are sensitive to transformational invariance over time, and, therefore, that maintaining maximum transformational invariance will entail following the maximum output of these analyzers. Doing so has the effect of maintaining object constancy over the transformation. For example, when a square begins to rotate, it is perceived as such, not as a square that changes in perceived shape until it becomes a diamond and then changes back into a square again. The latter is what would be expected if the rotation were not followed by the reference frame initially used to code its shape, but were fixed in the same unchanging orientation. I suspect that the latter is what happens when people are shown a tight spiral pattern rotating, yet see it as circles contracting into the center.

Organizational Phenomena Revisited

We now begin to see how motion can be analyzed and constancy can be maintained within such a transformationally based system. The transformations are initially coded by the higher order analyzers and then used to achieve and maintain maximal constancy. The motion finally perceived is not a simple function of the motion analyzers, since it too depends on a reference frame that maximizes invariance. This is accomplished by an attentional mechanism that finds and follows maximal output levels within the analyzers sensitive to motion and their higher-order analyzers.

Transformations of this reference frame can compensate for stimulus transformations, thereby maintaining constancy. It seems necessary that much of this must be done outside conscious attention, however, because there hardly would be enough of it to go around. More likely, once an object's representation has been established in memory, its representational schema can follow the approriate reference frame without conscious attention. This monitoring process

would only require conscious attention if something unexpected were to occur, such as the object disappearing or changing its intrinsic properties.

Many perceptual grouping phenomena are a natural result of this attentional process working within a transformationally structured space. It seeks the maximal amount of transformational invariance at different levels of resolution and codes elements together that are closely related within the analyzer space. Given that attention can only cover a portion of the analyzer space and that it is attracted to local maxima, it will tend to code together items that are transformationally similar.

Finally, reference frame effects result from an attentional mechanism that is centered on a position within the analyzer space and a coding scheme, more fully specified elsewhere (Palmer, in press), that descrites shape relative to the reference values of the frame. The fact that frame effects generally show that global structure affects local structure more strongly than vice versa suggests that global information tends to dominate in determining the position of the attentional frame. Higher order stroture of the whole configuration seems to strongly affect the placement of the reference frame and, therefore, to influence the resulting perception.

In all of these phenomena it is clear that the system's preference for invariance over transformations holds within the three dimensional space of the world. It may hold in the two dimensional space of images as well, but when the two conflict, the simpler three dimensional solution generally dominates. We have been discussing the analysis of two dimensional images, and it is not entriely clear how to extend the proposal into the third dimension. One possibility would be to add second, three dimensional level that embodied the same design features, but in a higher dimensionality. Another would be to translate the relevant simple transformations in three dimensions into their complex counterparts in two dimensions. I do not yet have a well defined proposal to make on this difficult issue.

The Importance of Systemic Structure

Before closing, I want to say a few words about an initeresting property of the theory quite apart from its ability (or lack thereof) to account for perceptual phenomena. I am intrigued by it systemic nature. The reader may have noticed that in explaining the theory I made almost no reference to the specific nature of the analyzers that comprise the pieces of the system. At one point I said that they might be something like bar- or edge-detectors, but that was merely for illustration. In fact, it makes very little difference what the analyzers look like as long as they satisfy certain symmetry conditions: namely, they cannot be symmetrical about any transformation proposed to exist within the analyzer space. The reason should be clear. If the analyzers are invariant over a transformation, then that transformation cannot exist as a dimension within the analyzer space. For example, if all the analyzers were rotationally symmetric, as are circular center-surround receptive fields, then the system could not support the orientation dimension or the higher order rotational motion analyzers.

In any case, the fact that the constraints on the system are so weak suggests that it is primarily the structure of the whole system that is doing the work. Indeed, this must be true if the basic building blocks of the system are transformational relations. Transformational relations are an emergent property of systems of analyzers; they are simply undefined for any individual analyzer without a systemic context of other analyzers. This suggests that the internal structure of individual analyzers might best be considered in terms of their functional role within the system. Further, the nature of the elements of the system might actually be determined by optimization of their functional roles within the system as a whole (Palmer, in press). I think these are interesting and important notions for perceptual theory. They hark back to the gestalt claim that emergent properties of whole systems play the critical role in understanding perceptual phenomena. Perhaps they were right.

REFERENCES

Attneave, F. Some informational aspects of visual perception. Psycholclogical Review, 1954, 61, 183-193.

Attneave, F. Triangles as ambiguous figures. American Journal of Psycholcgy, 1968, 81, 447-453.

Cutting, J. E. Coding theory adapted to gait perception. Journal of Experimental Psychology: Human perception & Performance, 1981, 7, 71-87.

Dunker, K. Ueber induzierte Bewegung. Psychologishe Forschung, 1929, 12.

Farrell, J. E., & Shepard, R. N. Shape, orientation, and apparent rotational motion. Journal of Experimental Psychology: Human Perception & Performance, 1981, 7, 477-488.

Garner, W. R. The processing of information and structure. Potomac, Md.: Erlbaum, 1974.

Hubel, L. H., & Wiesel, T. N. Receptive fields, binocular interaction, and the functional architecture of the cat's visual cortex. Journal of Physiology (London), 1962, 160, 106-154.

Johansson, G. Configuration in event perception. Stockholm: Almqvist & Wiksell, 1950.

Johansson, G. Visual perception of biological motion and a model for its analysis. Perception and Psychophysics, 1973, 14, 201-211.

Kosslyn, S. M. Image and mind. Cambridge. Mass.: Harvard University Press, 1981.

Leeuwenberg, E. L. J. A perceptual coding language for visual and auditory patterns. American Journal of Psychology, 1971, 84, 307-350.

Navon, D. Forest before trees: The precedence of global features in visual perception. Cognitive Psychology, 1977, 9, 353-383.

Palmer, S. E. Visual perception and world knowledge: Notes on a model of sensory-cognitive interaction. In D. A. Norman & D. E. Rumelhart (Eds.), Explorations in cognition. Hillsdale, N.J.: Erlbaum, 1975.

Palmer, S. E. Hierarchical structure in perceptual representation. Cognitive Psychology, 1977, 9, 441-474.

Palmer, S. E. What makes triangles point: Local and global effects in configurations of ambiguous triangles. Cognitive Psychology, 1980, 12, 285-305.

Palmer, S. E. Symmetry, transformation, and the strcture of perceptual systems. In J. Beck & F. Metelli (Eds.), Representation and organization in perception. Hillsdale, N.J.: Erlbaum, in press.

Palmer, S. E. Reference frames in shape perception. In preparation.

Palmer, S. E., & Bucher, N. M. Configural effects in perceived pointing of ambiguous traingles. Journal of Experimental Psychology: Human Perception & Performance, 1981, 7, 88-114.

Rock, I. Orientation and form. New York: Academic Press, 1973.

Shepard, R. N. & Judd, S. A. Perceptual illusion of rotation of three dimensional objects. Science, 1976, 191, 952-954.

Wallach H., & O'Connell, D.N. The kinetic depth effect. Journal of Experimental Psychology, 1953, 45, 205-217.

Wertheimer, M. Untersuchungen zur Lehre von der Gestalt. Psychologishe Forschung, 1923, 4, 301-350. Translated in W. D. Ellis (ed.), A sourcebook of gestalt psychology. New York: Harcourt, Brace & Co., 1938.

Weyl, H. Symmetry. Princeton, N.J.: Princeton University Press, 1952.

Rational Processes in Perception

Alan Gilchrist and Irvin Rock

In this paper we will give our reasons for believing that certain current attempts to explain perceptual phenomena on a lower level in terms of known sensory mechanisms are untenable. We will do this by focussing on two topics, lightness perception and the perception of apparent motion. We will summarize some older data (not all of which are sufficiently known) and will describe some recent work of our own. Finally, on a more positive note, we will try to indicate the direction that a theory must take if it is to deal effectively with these phenomena.

Lightness Perception

We will begin with the assumption that Helmholtz was essentially wrong in his belief that an object's lightness can be inferred by interpreting the luminance reflected by it to the eye in terms of the amount of illumination falling on it. Such a process requires unequivocal information about the illumination whereas the only information directly available is the intensity of light, or luminance, reflected by each surface in the field. Each such luminance is the joint product of the reflectance property of the surface and the illumination falling on that surface. Rather we will assume that the perceived shade of gray of a surface is governed primarily by the luminance of that surface relative to the luminance of neighboring surfaces as Hering (1920) suggested and as Wallach(1948) elegantly demonstrated. There is now fairly wide agreement among investigators on this general principle.

But what is the underlying explanation of it? There is great appeal in Hering's suggestion of reciprocal interaction, i.e. that a bright region of the field would have a darkening effect on an adjacent region and a dark region would have a brightening effect on an adjacent region. We now know for a fact that the rate of discharge in one nerve fiber is attenuated when a neighboring fiber is stimulated by light. Thus such lateral inhibition can plausibly be invoked to explain why the apparent lightness of one region is governed by the extent of stimulation of an adjacent region (see Cornsweet, 1970; Jameson and Hurvich, 1964).

In fact, given lateral inhibition as a known sensory effect, one might have been able to predict the phenomenon of contrast, even if it had never been observed (although questions can be raised about the spatial distance over which such a mechanism can be expected to occur). Surrounding a gray region by a white one should lead to diminished discharging of retinal fibers stimulated by the gray region; surrounding another gray region of the same value by a black one should lead to increased discharging of retinal fibers stimulated by that gray region because of a release of inhibition. Thus one of these gray regions should look lighter than the other and so it does. An implicit assumption here is that the phenomenal shade of gray perceived in a given region is a direct function of the rate of discharging of fibers stimulated by that region.

The fact of constancy of lightness can be explained along the same lines. When the illumination falling on a surface changes, then the luminance of all adjacent regions rises and falls together. Thus, while the rate of discharging of cells stimulated by a gray region should increase when illumination increases, so too should the rate of discharging of cells from a surrounding white region increase. The latter will increase the inhibition on the former with the net result of little if any change in the absolute rate of discharging of those cells. Therefore the perceived lightness should remain more or less constant and so it does. Note again, however, the assumption, here explicit, that the phenomenal lightness is a direct function of the rate of discharge of the appropriate fibers.

Underlying this assumption is another assumption about how the visual system works that Gilchrist (1981) has called the photometer metaphor. Just as the signal produced by a photometer is a direct function of the light falling upon it, so the perceived lightness of each point in the field is assumed to be a direct function of the rate of discharging of the cells stimulated by each such point. With the knowledge that has been available about light, about the formation of the retinal image, and about photochemical processes and nerve physiology it is understandable why such a view has become so deeply ingrained as not even to be explicitly recognized as an assumption. Given this assumption, phenomena such as contrast and constancy, in which lightness does not correlate with luminance, seem to require an explanation long the lines of lateral inhibition.

There is now, however, reason to reject this approach. Evidence has been accumulating to support the theory that the perception of lightness (and chromatic color) is based on information at the edges between regions of differing luminance (or hue). Homogeneous regions between edges are then "assumed" to have the lightness or color indicated by these edges. There is overwhelming evidence (Yarbus, 1967; Whittle and Challands, 1969; J. Walraven, 1976) that the visual system responds to changes in stimulation, not to an unchanging state of stimulation. This is normally guaranteed by continuous eye movements, for vision. Whenever an image can be held stationary on the retina for a few seconds, all visual experience stops. These facts are inconsistent with the photometer metaphor and they strongly indicate the crucial nature of edges or gradients in the retinal image since this is where stimulation changes in the normal moving eye. Krauskopf (1963) has shown that when the boundary of a surface is prevented from moving on the retina, its color will disappear and be replaced by the color of the surrounding region, signalled by the boundary of that region.

It seems unlikely that any absolute luminance information would be picked up in this way and yet it now seems quite possible that the visual system achieves what it does using only relative information. Even a simple edge-relations approach goes a long way toward explaining lightness constancy since the luminance ratio between two adjacent surface colors remains the same even when illumination changes.

The important point here is that there is no need to invoke a concept such as lateral inhibition to explain constancy. Once the photometer assumption is made explicit and in fact, is displaced by the concept of edge information, the whole edifice collapses. Of course lateral inhibition is a well-established physiological fact. It is probably part of the process whereby the ratio at an edge is determined. But we don't believe that lateral inhibition solves any of the basic problems of constancy. The concept of an exaggeration or enhancement of edge ratios seems unnecessary and illogical. If lateral inhibition exaggerated an edge ratio, it would do so in the same way every time an edge of the same value were present on the retina. A given edge ratio would be specified by a given neural signal, with or without an exaggeration function. Therefore the exaggeration function doesn't seem to add anything of explanatory value.

Certain problems turn out to be dissolvable pseudo-problems with the adoption of an edge-relations approach. One of these is constancy under changing illumination. On the other hand, other problems emerge, although they are more tractable. For example, how do we now explain the constancy of surface lightness as the surface is viewed against differing backgrounds? The luminance ratio at the edge of a surface can change

dramatically as it is placed on different backgrounds and yet lightness perception remains almost unchanged.

For example, in the classic example of lightness contrast, the gray square on the white background has an edge ratio that is radically different (even opposite in sign) from that of the gray square on the black background. Thus, under a simple edge theory they ought to appear radically different in lightness, and yet they appear almost the same. This suggests that lightness is not determined simply by the boundary of a surface, but by the relationship between that boundary and other boundaries. Presumably the boundaries of the squares themselves only signal departures from a background lightness, which in turn is signalled by the boundary of each background. Thus the edge dividing the white and black backgrounds signals the relationship between the two background lightnesses and we might expect that this edge will be as critical to the lightness of the targets as the edges of the targets themselves. In fact Gilchrist and Piantineda (unpublished experiment) have found that if that edge is retinally stabilized, the two gray squares turn black and white respectively, just what we would expect based simply on the ratios at the edges of the gray squares. We might say that the assignment of lightnesses to the various regions is the end result of a computational process in which information from all edges present is integrated. Arend (1973) and Land and McCann (1971) have proposed similar schemes.

If such computational processes occur and are governed by remote as well as local edge information, the reader may well wonder about the achievement of constancy. Consider the typical case where two gray disks of equal reflectance on the same background are unequally illuminated because one region and its immediate background are in shadow. Earlier we said that constancy could be explained on the basis of the equal ratio of each gray region to its background. That would be true in the example under consideration. But now we have also said that the presence of other edges enters into the equation. The shadow edge can easily have a ratio as great as a white-black edge and, more probably, even greater. If this is entered into the computation, constancy would fail; the disks would be seen as different shades of gray in accordance with the luminance difference between them. The equal disk-to-surround luminance ratios here logically cannot signify that the two grays are equal if the gray regions are seen as on backgrounds of <u>different</u> luminance values, or so it would seem.

Unless the perceptual system can discriminate between reflectance edges and illumination edges, that is between changes in the pigment of the surface and changes in the amount of illumination shining on the surface. If so, perhaps illumination edges would not be included in the computation of surface lightness values. There is now strong evidence of just such discrimination of reflectance and illumination edges (Gilchrist, in press). If observers view the two disks under the conditions just described, they typically do perceive the two grays as almost equal, i.e. constancy is achieved. Moreover, they perceive both sides of the background as white with one side in shadow. Thus the central edge is apparently correctly identified as an illumination edge. If, however, the observers view the display through an aperture that permits only part of the background and the two gray regions to be seen, and if the edge of the shadow is reasonably sharp, the grays no longer look equal, constancy is destroyed. Moreover, the observers now perceive the two sides of the background as unequal in lightness. Thus the edge is interpreted as separating different reflectances, not different illuminations. In this condition only then does the central edge enter into the process of computing the lightness of the gray disks.

We reported earlier that if the boundary between the black and white backgrounds in the traditional contrast pattern is made to disappear through retinal stabilization, one gray square turns black and the other turns white. This provides some of the best evidence for the concept of edge integration. A similar experiment was done by Gilchrist, et. al. (in press) that demonstrates the importance of the distinction between reflectance edges and illumination edges. When the boundary between the white and black backgrounds is made to look like the edge of a shadow, the two squares will also turn black and white respectively, just as in the stabilized-image experiment. Thus when an edge is identified as an illumination edge, it seems to drop out of the integration process for surface lightness just as if it were invisible.

It would take us too far afield to enter into a <u>full</u> discussion of precisely how the perceptual system discriminates illumination from reflectance edges. While the presence of penumbra at an illumination edge may be one source of information it is not the only one and is not necessary in the experiments just described.

Before discussing the important role that depth perception plays in discriminating edges, it is worth considering the ramifications of what we have just discussed for the notion of lateral inhibition or any other theory of neural interaction which seeks to explain the important effects of remote edges on what is perceived in regions adjacent to other edges. For now in addition to other difficulties such a theory faces, in dealing with such "remote" effects it would have to be argued that such effect <u>do not occur at all</u> when those remote edges are interpreted as representing illumination rather than reflectance differences. In fact, it is interesting to note that apparently no one has noticed that when contrast effects are applied to illumination edges, they not only fail to result in constancy, but they actually <u>make matters worse</u>. Constancy requires that we explain how the perception of surface lightness could be the same on both sides of an illumination edge, given the difference in luminance. Applying a mechanism here that further exaggerates the luminance difference is a little like bringing water to a drowning man.

Recent experiments (Gilchrist, 1977, 1981) have demonstrated the role of depth perception in distinguishing an illumination edge from a reflectance edge. In one experiment an artificial interposition cue was used to make a target square appear as located either in a near plane, dimly illuminated, or in a far plane, brightly illuminated, as shown in Figure 1. In terms of the retinal image, the target square was always flanked by a surface of much lower luminance to its lower right and by a surface of much higher luminance to its upper left (note relative luminances given in Figure 1). In space, however, the low luminance surface was located in the near plane while the high luminance surface was located in the far plane. The results show that lightness is determined by the luminance ratio of the target square to its coplanar neighbor. Thus the target square looked white when it appeared in the near plane but black when it appeared in the far plane. In other terms, the edge dividing the target from its coplanar neighbor was treated as a reflectance difference while the edge dividing the target from its non-coplanar neighbor was presumably treated as representing an illumination difference. Since the retinal image was essentially the same in both conditions, this result is inconsistent with an explanation based on lateral inhibition.

In another experiment involving planes meeting to form a dihedral angle, these ideas were put to a more rigorous test in which predictions based on perceived coplanarity would be the opposite of predictions based solely on retinal ratios.

The experimental arrangements are shown in Figure 2. In the horizontal plane, a black target tab extended out into space from a larger white square. In the vertical plane, a white target tab extended upward into space from a larger black square. Thus each target

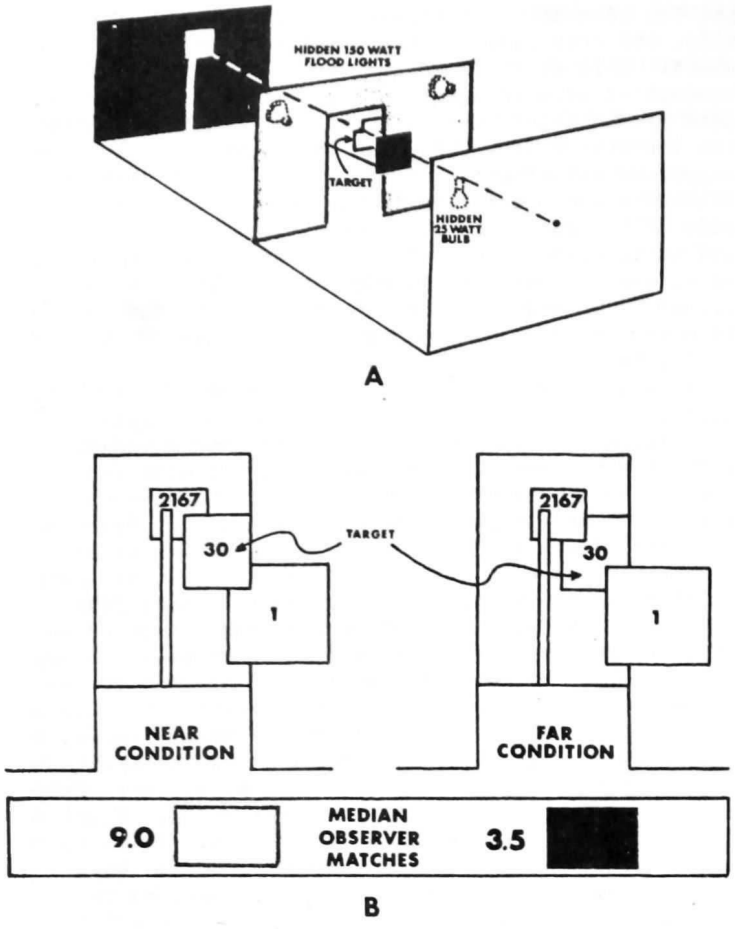


Figure 1

tab was seen against the background square that was in a separate plane. The horizontal surfaces received about 30 times as much illumination as the vertical surfaces, or just enough illumination difference to make the luminance of the black target tab equal to that of the white target tab. Given the viewing perspective of the observer, 45 degrees from each plane, the display was similar to traditional contrast displays; two targets of equal luminance on bright and dark backgrounds respectively. Thus a theory based on lateral inhibition would clearly predict that the target on the bright background, in this case the upper target, should appear darker than the other target, although the exact magnitude of the effect is harder to determine. On the other hand, if lightness is really based on luminance relationships within planes, then each tab should be compared with the larger background square that lies in the same plane, even though it is adjacent only along one edge of the tab. Thus not only would the coplanar ratio principle predict that the upper tab would appear lighter, not darker, than the lower tab, it would predict that the upper tab should look white and the lower tab black.

In fact the latter result was actually obtained. Figure 2 shows the median Munsell matches (next to samples of those Munsell values) obtained from naive observers. Moreover, since the target tabs were actually trapezoidal in shape, they could be made to switch perceived planes when viewed monocularly. In that condition of the experiment the perceived lightnesses of the tabs also switched, with the lower tab now appearing white and the upper tab appearing black. Since these changes in perceived lightness were produced solely by a change in depth perception, with no change in the retinal image, these data raise difficulties that may be insurmountable for current theories based on lateral inhibition.

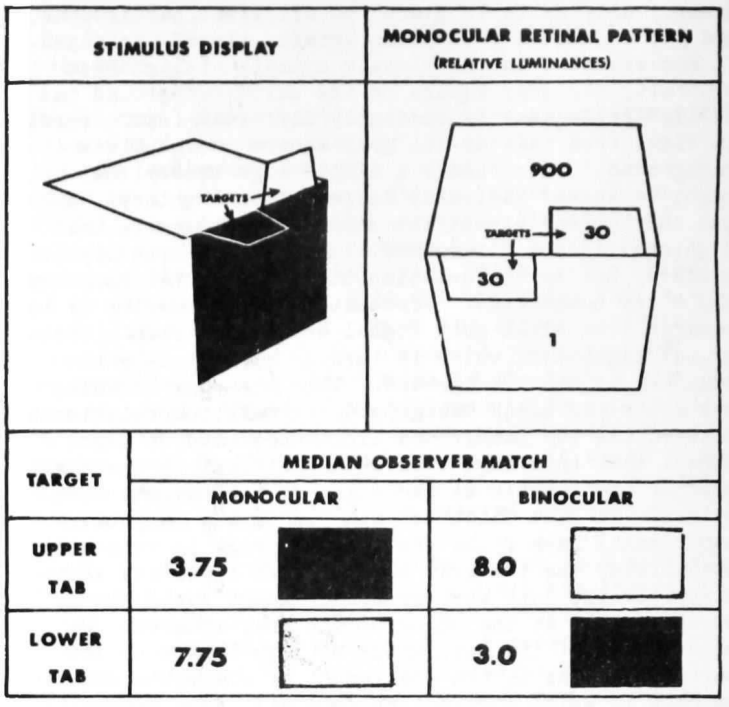


Figure 2

Apparent Motion

Although we know a good deal about the conditions that produce the illusory impression of motion referred to as apparent motion, we still do not understand why it occurs or, for that matter, why it only occurs under certain conditions. What we know about this effect is that given the sudden appearance of object a, its sudden disappearance, followed typically by just the right time interval of object b in just the right new spatial location, one tends to see motion of a to b. The currently favored explanation is that a motion-detector cell in the brain will discharge even if the appropriate receptor field of the retina is stimulated discontinuously by two points rather than by a point moving over the retina. Such cells do seem to exist in various species of animals (Grüsser-Cornehls, 1968; Barlow and Levick, 1965).

However, the fact is that it is not necessarily the case that the conditions for apparent motion perception entail stimulation of separate retinal regions. Ordinarily that is the case, since a and b are in separate spatial locations and the eye is more or less stationary. What seems to matter is the perception of a and b in separate locations in space.

To get at this question an experiment was performed in which the observers had to quickly move their eyes back and forth synchronous with the onset of a and b so that each stimulated the same central region of the retina, rather than as, more typically, two discretely different loci (Rock and Ebenholtz, 1962). Therefore, the conditions for apparent motion might be thought not to exist. Yet, the observer does locate a and b in phenomenally discrete places in the environment. The result was that although nothing was said to the observers about motion that might create an expectation of perceiving motion, most of them nonetheless spontaneously did. This experiment seems to prove that, in humans at least, it is not necessary to explain stroboscopic motion in terms of a sensory mechanism that detects sudden change of retinal location. There is neither change of retinal nor cortical locus of projection of a and b here.

An entirely different view that has been presented by Rock (1975) is that the impression of motion is a solution to the problem posed by the rather unusual stimulus sequence. First a inexplicably disappears. Then b inexplicably appears elsewhere. By "inexplicable" we

mean that when an object in the world disappears as we are looking at it, it is generally because another object moves in front of it or it is occluded by another object because of our motion. However, when a stationary object suddenly and rapidly moves to another location, it does tend to disappear from one location and to appear in another. Therefore, perhaps this state of affairs in a stroboscopic display suggests the solution of motion.

Given that potential solution, the question arises as to whether it is acceptable. Motion from a to b does account for the brief stimulation by a and b, but isn't the absence of any visible object between the locus a and b a violation of the requirement that a solution be supported by what is present in the stimulus? If the solution is "a moving across space to b" doesn't this call for stimulus support in the form of continuously visible motion across that spatial interval? Ordinarily that would be true, but it is a fact that has been demonstrated that for very rapid motion of an actually displacing object, little more than a blur can be seen in the region between the terminal locations (Kaufman et al, 1971). In fact it was shown that if the terminal locations are occluded, no motion of a moving object is seen. Therefore when the spatial and temporal intervals between a and b in a stroboscopic display are such as would correspond with the real motion of a rapidly displacing object, the absence of continuously visible movement need not act as a constraint against perceiving movement. In fact, this analysis may explain why slow rates of alternation do not lead to the impressions of motion. By "slow rate" we mean a condition with a relatively long interval between the disappearance (offset) of a and the appearance (onset) of b. Such a rate would imply a slowly moving object and a slowly moving object would normally be seen throughout the spatial interval between a and b. Therefore the absence of object motion over that interval at slow rates of alternation is a violation of the requirement of stimulus support. Hence the movement solution is not acceptable at slow rates even if the offset and onset tend to suggest this solution.

While on this topic of rate we might briefly comment on the case where the alternation is very rapid, i.e. a zero or only a minimum interval between the offset of a and onset of b. If the "on" time of a and b is itself very brief, this state of affairs will result in a and b being visible simultaneously by virtue of neural persistence. But if a is visible when b appears, the solution that a has moved to b is not supported or one might say, is contradicted. This deduction was tested by using rates of alternation that ordinarily do produce the stroboscopic effect but with the following variation: First a appears, followed by the usual blank interval; when b appears so does a (in its original location). Therefore the sequence of events is: a; a and b,: a; a and b, etc. If the presence of a during the exposure of b violates the requirements of the motion solution, then observers should not achieve a stroboscopic effect under these conditions. Our observers did not. If, however, the display is changed so that the a object that appears concurrently with b but in the same location as a is somewhat different than the a that appears alone, observers do perceive a moving to b. The sequence is a; a' and b,: a; a' and b; etc.

It was noted above that in the typical experiment on stroboscopic motion, a and b inexplicably disappear and appear. What was meant was that no rationale is provided to the observer of why they appear and disappear such as is the case when things in the environment suddenly appear or disappear because another object in front suddenly moves out of the way or in the way. This suggested the following kind of experiment. Suppose we cause the retina to be stimulated by a and b in just the right places at just the right tempo, etc., but by a method in which we move an opaque object back and forth, alternately covering and uncovering a and b.

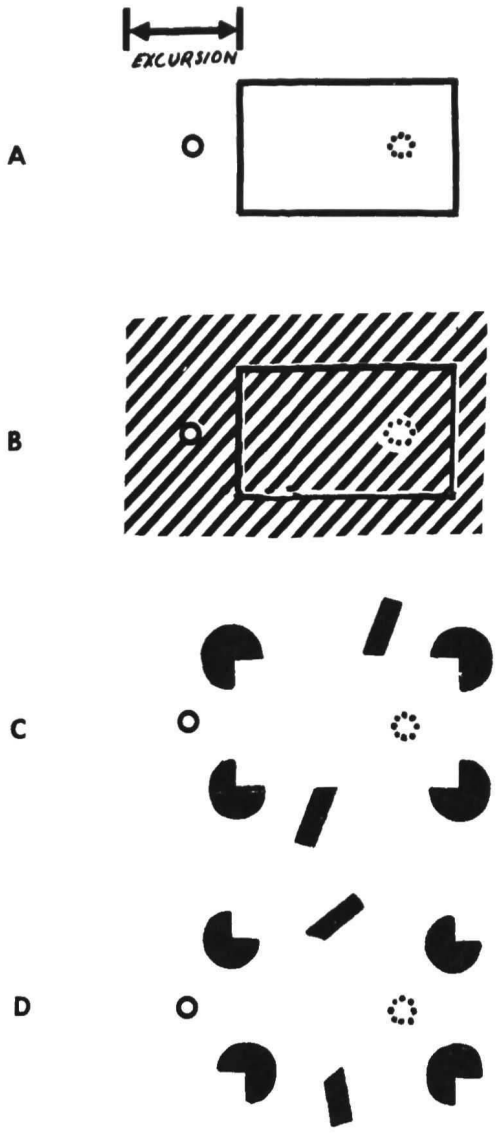


Figure 3

(See Figure 3B) As far as the sensory theory of apparent motion is concerned there is no obvious reason why these conditions should not produce an impression of a and b moving. But from the standpoint of problem solving theory, we have now provided an explicable basis for the alternate appearance and disappearance of a and b, namely, that they are there all the time but undergoing covering and uncovering. Therefore the perceptual system may prefer this solution or at least we are offering it a viable alternative not usually available (see Stoper, 1964; Sigman and Rock, 1974).

The subjects rarely perceived motion of the dots here. Some may object that the presence of the actually moving rectangle interfered in some way with perceiving stroboscopic motion. It is, after all, an unusual, atypical, way of studying such motion. The rectangle may draw the subjects' attention or otherwise inhibit motion perception of the dots. For this reason a slight change was introduced, one that had another purpose to it as well. Suppose the rectangle moves, but a bit too far, far enough no longer to be in front of where the dot had been. But by a method, the details of which need not be discussed here, things were so arranged that when the rectangle is in its terminal location, the dot is nonetheless not visible.

Now it is no longer a fitting or intelligent solution to perceive a and b as two permanently present dots that are simply undergoing covering and uncovering. For it can be seen that in fact the rectangle is not covering the spot in its terminal location and yet the spot is not visible (violation of the stimulus-support requirement). Therefore the best solution is again one of

movement and that is what the subjects perceived. Note that this experiment serves as a control for the objection raised to the first one; the moving rectangle here does not interfere with perceiving motion of the dots.

Another variation performed is based on the idea that for the covering-uncovering solution to be viable, the covering object must appear to be opaque. If it does not, it can hardly be covering anything. This factor was manipulated in an experiment illustrated in Figure 3B. The actual stimulus conditions are very similar but in one case, because the oblique lines within the rectangle are stationary and aligned with all the others, the rectangle looks like a hollow wire perimeter. In a control condition the lines inside it moved with the rectangle, and it looked like an opaque object. The difference in results is very clear: when the rectangle appeared to not be opaque, subjects by and large perceived movement whereas in the case of the opaque-appearing rectangle, they did not. A hollow rectangle is in contradiction of the property of opacity required by the covering-uncovering solution.

In a final experiment, conditions were such that no physical contours at all moved back and forth in front of the dots. There was, however, a phenomenally opaque object that moved, one based on illusory contours, as illustrated in Figure 3C. The great majority of subjects did not perceive movement. In a control experiment, illustrated in Figure 3D, the orientation of the corner fragments was changed so that no subjective rectangle was perceived and this array was moved back and forth. Now the majority of subjects did perceive movement.

It should be noted that in all these cases where a covering-uncovering effect is perceived there is no reason why movement of the dots could not have been perceived as well. That is to say, if the observer were to see an opaque rectangle moving back and forth and, simultaneous with this, a dot stroboscopically moving in the opposite direction, such a solution would also account for the stimulus sequence. Conversely everything implied by that solution is represented in the stimulus, and no contradictory perception is occurring. Therefore the tendency to perceive dots undergoing occlusion and disocclusion rather than dots moving, represents a preference for one solution over the other. The preferred solution is obviously related to a very basic characteristic of perception, namely, object permanence, the tendency to assume the continued presence or existence of an object even when it is momentarily not visible for one reason or another. But given the very strong predilection we have to perceive apparent motion even under the most unlikely conditions, it remains a problem as to why it is not perceived in this situation and the object-permanence solution is preferred. A possible answer is that the covering-uncovering solution accounts for all stimulus change by one "cause": a moving rectangle covering and uncovering spots that are continuously present. The other solution entails two independent events that are coincidentally and unaccountably correlated; a rectangle moving in one direction and in anti-phase to spots moving in the opposite direction.

There is another line of evidence that also strongly supports a problem-solving interpretation of stroboscopic motion. If the stimulus consists of more than a single dot or line, the problem arises of what in a is seen moving to what in b. To make the point clear, suppose that a and b each consist of a two-by-three matrix of dots. What will be seen here is the rectangular grouping moving as a whole (Ternus, 1926). Apparently the perceptual system seeks a movement solution that will do justice to the object as a whole. Indeed, were this not the case, the motion perceived in moving pictures would be quite chaotic, because it is typically objects consisting of many parts that change location from frame to frame (and often many such objects are simultaneously changing locations in either the same or varying directions). Yet this outcome is not predictable at all in terms of the other kinds of sensory theories mentioned earlier.

A related example is the perception of motion of complex stimuli such as the line drawings of three-dimensional cuboid figures that Shepard and his associates have used in the mental rotation studies. Shepard and Judd (1976) presented two perspectives of such figures in a stroboscopic motion paradigm and showed that, at the appropriate rate of alternation, observers perceive these objects rotating through the angle necessary to account for the change in perspective from a to b. This effect clearly implies that the perceptual system deals with the problem of accounting for the differences in a and b by an intelligent motion solution. A further finding of interest is that the optimum rate of alternation for achieving a continuous coherent rotation of a rigid whole object was an inverse function of the angular difference as implied by the two perspectives views. In other words, the greater the angle through which rotational motion was seen, the slower the rate of alternation had to be.

This finding can be considered to be in keeping with one of Korte's Laws which states that optimum apparent motion is preserved when the spatial separation between presentations of a and b is increased by increasing the time interval between presentation of a and b. This law makes sense if one assumes that the perceived speed of rotation is constant. If therefore the mental representation of the object has to rotate through a greater angle, more time is required.

Further support for this interpretation is provided by an experiment which asked the following question: Is it the retinal spatial separation or the perceived spatial separation that governs Korte's Law? Perceived separation was varied by creating conditions in which a and b appeared at differing distance but were always located so as to project to the eye in the same retinal loci (Corbin, 1942; Attneave and Block, 1973). The experiments demonstrated that it was the perceived spatial separation, not the retinal separation that enters into Korte's Law.

A problem-solving theory can account for these facts. It offers an explanation of why motion is seen. Unlike other theories, it takes as a point of departure and is quite compatible with the fact that the conditions leading to motion perception entail change of perceived location rather than change of retinal location. It offers a rationale for the known facts about alternation, i.e. why movement is perceived only within a certain range of middle values of inter-stimulus interval. It can deal easily with the kinds of perceived transformations or movements that occur when a and b are more than single dots or lines, such as groupings or forms with sub-parts, or complex three-dimensional figures in differing orientations. Finally it permits us to predict instances where no motion will be perceived despite the maintenance of the spatial and temporal parameters that ordinarily produce the stroboscopic effect.

On the other hand this theory does not as yet explain all the known facts. It does not explain the reported findings that motion is seen more readily if both a and b are placed so that their projections fall within one hemisphere of the brain (Gengerelli, 1948); nor does it explain why the effect is more readily obtained if a and b stimulate one eye compared to the case where a stimulates one eye and b the other (Ammons and Weitz, 1951). However, these findings have never been replicated and warrant careful re-examination. And finally a problem-solving theory might be considered to be inappropriate as an explanation of the stroboscopic effect that seems to occur in decorticated guinea pigs (Smith, 1940) or newly born lower organisms such as

fish or insects (Rock, Tauber, and Heller, 1965).

However there now seems to be fairly good evidence that there are two kinds of apparent motion (Broddick, 1974; Anstis, 1980). One kind, referred to as the short-range process, occurs over very small angular separations of a and b. There is reason for believing that this kind may be based on motion-detector neurons responsive to a small shift in stimulation on the retina. The other kind, referred to as the long-range process, occurs over larger angular separation of a and b. This process is probably not based on the activation of motion-detector neurons. Most if not all of the evidence discussed above pertains to this long-range process. The short-range process thus seems to have a direct sensory basis whereas the long-range process seems to have a cognitive basis. In the light of this distinction, it is possible that the findings referred to in the previous paragraph are explicable in terms of the short-range process.

Conclusion

At this point we should step back from these empirical studies and see what general lessons can be drawn as to the nature of theories of perception. If the visual system is to achieve a faithful representation of the physical world then the organization of its own processes must somehow mirror the organization of the world. Any theory of perception that does not take this point into account will ultimately fail.

In certain theories of perception, constancy and veridicality are fortuitous outcomes that occur only under some circumstances. This is not good enough. Both the logic of what the perceptual system must accomplish and the emperical evidence of what it does achieve demand a theory in which constancy is inevitable, not accidental.

Herein lies the danger of theories based on simple and limited physiological findings. Unless the physiological finding can be seen as part of a larger process that "homes-in" on reality in an inevitable way, that physiological finding is likely to be misunderstood. This is the problem with viewing lateral inhibition as an exaggeration or distortion process. If we cling to the photometer metaphor, to the assumption that fundamentally the visual system measures the intensity (and perhaps the wavelength) of the light at each point in the image, then it is not surprising that some kind of distortion process will be required to transform the array of photometer readings into something vaguely representing visual experience.

There is no need to talk as though the intensity of light at point A in the image is "affected" by the intensity of light at point B. The fact is that the light at point A, seen by itself, would be perceptually meaningless. Having a second intensity of light present in the visual field doesn't merely change the first amount of light, it literally establishes a relationship and the apprehension of this relationship produces lightness perception in its simplest form.

Contrast theories are usually thought to be relational theories, but they are not. As Koffka (1935) has correctly pointed out, the ultimate correlate of lightness perception in a contrast theory is still an absolute amount of light, not a relationship. The contrast process only allows the absolute value of one region of the field to be changed as a function of other values. But it is still that absolute value that reigns. And the reason that it has to be changed is that as an absolute value, it will always be out of touch with visual experience, which involves relationships. In fact the history of theories of lightness perception is the history of different correction factors designed to bring the local luminance into correlation with perceived lightness. This has never worked and it is time we recognize that no theory based on absolute amounts of light can work. What is constant about a white surface, for instance, is its relationship to the rest of its environment.

It is not surprising that the visual system gets its critical information from the edge. This is the point in the image where the relationship between two amounts of light is represented. In more complex scenes each local edge relationship, or ratio, has to be seen in relation to other edge ratios. The concept of edge integration that we have discussed does not involve any distortion or exaggeration process. Rather it involves the proper organizing of certain local relationships in order to make explicit a more global relationship that was only implicit in the local relationships. At the level of cognition the same function is served formally by the syllogism.

Fundamentally the visual system must be logical because the world is logical. The world is not put together in a random or capricious way. If a rectangular object is incapable of obscuring a set of diagonal lines it will also be incapable of obscuring a luminous spot. How inefficient it would be if local percepts were allowed to coexist with other local but contradictory percepts. A system that excludes contradictions from its global relationships has the tremendous advantage of reducing the ambiguity of its local relationships.

Perception and cognition seem to share this quality of excluding contraditions within their own domains. Of the two, however, perception seems to be the more successful. Of course it is possible to construct figures such as impossible triangles, or Escher drawings, which surprise us by the extent to which visual contradictions are tolerated. But it is the rareness of such visual contradictions that leads to our delight at such figures. Examples in which the cognitive system fails to exclude contradictions are unfortunately too numerous to mention. One only needs to turn to political speeches, or the Bible, or journal articles to find a gold mine of examples.

Seeing, then, is like thinking, at least in many of its formal properties, and this may be because thinking is like seeing. That is, seeing may be the primitive form of thinking, the basic prototypical form that shows how relationships are to be integrated in order to correctly represent the world. Seeing, of course, had to come first, and it must be there even in mosquitoes. Thinking, however, allows us to integrate relationships that extend beyond the time and space limitations of the visual system.

Perhaps the world looks the way it looks because we are what we are physiologically. But it should not be forgotten that the world existed before we did and thus, as we learn from evolution, we are what we are because the world is what it is.

THE ROLE OF SPATIAL WORKING MEMORY IN SHAPE PERCEPTION

Geoffrey E. Hinton

MRC Applied Psychology Unit
Cambridge, England

ABSTRACT

Three demonstrations are presented and used to support a number of apparently unrelated claims about the internal representations that people have when they perceive or imagine a spatial structure. The first demonstration illustrates properties of the spatial working memory that enables us to integrate successive glimpses of parts of an object into a coherent whole. The second demonstration shows that our ability to generate a mental image is severely limited by the form of our knowledge of the shape of an object. The third shows that the shape representation which we create when we attend to a whole object does not involve creating the kinds of shape representations for the parts of the object that we would form if we attended to them and saw them as wholes in their own right. The real motivation for this medley of demonstrations and for the interpretations offered is that these phenomena can all be seen as manifestations of a particular kind of parallel mechanism which is described briefly in the last section.

I PERCEPTION THROUGH A PEEPHOLE

Fig. 1 illustrates a phenomenon called anorthoscopic perception that occurs when people perceive an object one piece at a time through a slit or peephole (Hochberg, 1968). Under suitable conditions people report that they have a perceptual experience of the whole object. They somehow integrate a number of separately perceived pieces into a single Gestalt. This means that they must be storing internal records of their perceptions of the individual pieces. The simplest theory of anorthoscopic perception is that the subject builds up an internal, picture-like representation part by part, and then uses this internal "picture" as a substitute for a retinal image in identifying the whole object. As we shall see, this theory has problems.

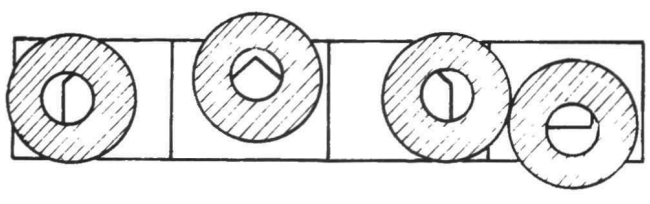


Figure 1. A cartoon strip showing a peephole moving around the outline of a shape. The fact that successive frames in the cartoon fall in different positions makes the task harder.

Retina-based versus scene-based frames

In the early stages of visual processing, the size, position and orientation of parts of the visual input are represented relative to the frame of reference defined by the retina. Anorthoscopic perception, however, cannot depend on storage in these early, "retina-based" representations because people typically fixate on the peephole, so all the different pieces of the object project to the same bit of the retina (Rock, 1981). Representations that encode the positions of the pieces relative to the retina would not allow us to perceive the whole object because the relative position of a piece within the whole is determined by where the peephole is, not by where the piece falls on the retina. It is just conceivable that as we move our eyes, the internal records of all the previously perceived pieces are correspondingly altered so that the records always encode where the piece is relative to the current retinal position, but this seems very unlikely.

What is needed is a way of representing where the pieces are that is not affected by eye-movements or even by movements of the whole person through space (Turvey, 1977). This can be achieved by using a temporary scene-based frame of reference that is defined by some larger contextual object or configuration within the external scene. If we keep a continually updated representation of the relationship between the retina and this scene-based frame, we can use it to convert from positions on the retina into positions relative to the scene before storage. These positions relative to the scene will be unaffected by subsequent eye or body movements. Obviously the scene-based frame will have to change from time to time, and it will have to have a scale that is appropriate to the scale of the parts we are attending to, but over a period of a second or two, perceptual integration of the results of successive fixations could be achieved by using a single scene-based frame of reference.

Post-categorical versus atomistic representations

In a picture-like representation, the shapes of objects are not explicitly represented -- it requires an interpretive process to extract them. Consider, for example, how a straight line is represented in an array. The line is decomposed into "atomic" fragments each of which is depicted by filling in one cell in the array. The absolute positions of the individual atomic fragments relative to the whole array are encoded directly and precisely, but there is no direct encoding of the straightness of the line, because this depends on the relative positions of the various fragments. Using this kind of atomic depiction it is impossible to represent the fact that a line is straight without representing precisely where it is relative to the whole array. It is impossible to be precise about shape and vague about position in a picture-like representation.

The memory used in anorthoscopic perception,

however, seems to allow just this combination of precision and vagueness. If a peephole is moved around a polygonal spiral (see Fig. 2) people often "perceive" a closed polygon. Their memory for the precise locations of the individual sides is poor and can be swayed by expectations about closed polygons, but they know that the sides are straight. This informal evidence that spatial working memory can be more precise about the shapes of pieces than about their positions implies that it contains explicit representations of shapes rather than being a picture-like collection of atomistic local features in which shapes are only implicit. A recent experiment supports this conclusion.

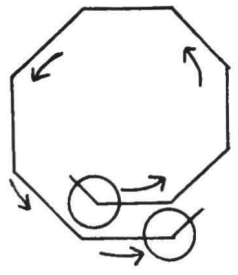


Figure 2. A peephole is moved around a polygonal spiral without revealing the free ends or the adjacent parallel sides.

Girgus, Gellman, and Hochberg (1981) have shown that it is considerably easier to "see" the shape of a whole object if the peephole is moved around the outline of the object than if the peephole jumps randomly from one part of the outline to another. The two different conditions were balanced so that the total exposure to any one part of the object was identical, so the contents of a picture-like store would be equally good in both cases. The obvious interpretation of this experiment is that when neighbouring parts of an object are exposed in succession, it is possible to form more complex chunks (shapes) and hence to reduce the number of chunks that must be stored in spatial working memory. When successive exposures are of widely separated pieces, either no chunks are formed, or chunks are created which do not correspond to the natural parsing of the whole object into parts. This type of explanation implies that the memory involved contains explicitly segmented and identified chunks.

II THE CUBE TASK

Hinton (1979) describes an apparently simple mental imagery task that people cannot do:

"Imagine a wire-frame cube resting on a tabletop with the front face directly in front of you and perpendicular to your line of sight. Imagine the long diagonal that goes from the bottom, front, left-hand corner to the top, back right-hand one. Now imagine the cube is reoriented so that this diagonal is vertical and the cube is resting on one corner. Place one fingertip about a foot above a tabletop and let this mark the position of the top corner on the diagonal. The corner on which the cube is resting is on the tabletop, vertically below your fingertip. With your other hand point to the spatial locations of the other corners of the cube."

It is fairly easy to imagine a cube in just about any orientation if the orientation is defined in terms of the natural axes of the cube. But when the diagonal is used to define the required orientation, we realise that relative to the diagonal, we have no clear idea where the various parts of the cube are. Our knowledge of the spatial dispositions of the parts of a cube is relative to the "intrinsic" frame of reference defined by the cube's own axes. Knowledge in this form is ideal for recognising the shape of a rigid object because whatever the object's actual size, position and orientation, the dispositions of its parts will always be the same relative to an intrinsic frame of reference based on the object itself (Palmer, 1975; Marr and Nishihara, 1978). So if the appropriate object-based frame can be imposed, the early retina-based representations which encode the positions of the parts relative to the retina can be recoded into object-based representations and this encoding will constitute a viewpoint-independent shape description that allows the object to be recognised.

I have now appealed to three different sorts of reference frame. The initial processing of the visual input uses representations relative to the retina; recognition of the shape of an object involves recoding these early retina-based representations into ones that are relative to an object-based frame; and anorthoscopic perception relies on storing the relationships of recognised shapes to a temporary scene-based frame.

III FRUITFACE

Fig. 3 shows a face composed entirely of pieces of fruit. Palmer (1975) reports that when subjects are shown this figure very briefly, they see it as a face without seeing the parts as fruit. The fruitface figure demonstrates that forming the Gestalt for a face does not depend on forming Gestalts for the parts. This is puzzling because to see the face we must form some representations of the parts and their relationships to the whole, since it is the relative dispositions of the parts within the whole that make it a face. One possibility which has not been much explored is that each part of the face can have two quite different internal representations. When the part is seen as a constituent of the face it receives a representation in which it is interpreted as filling the role of, say, an eye because of its crude overall shape and its relation to the whole face. When it is seen as a whole in its own right, however, it receives a quite different internal representation in which the rough shapes and dispositions of its parts cause it to be seen as a piece of fruit.

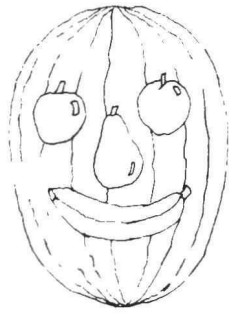


Figure 3. A face composed entirely of pieces of fruit. (After Palmer, 1975)

The idea that an object receives a quite different internal representation when it becomes the object of focal attention does not fit the popular view of attention as a kind of internal spotlight which can illuminate any one of a number of otherwise unconscious shape representations. However, the idea is very compatible with "early selection" theories (Triesman and Gelade, 1980) in which focal attention is constructive and is necessary for the generation of a shape representation.

The internal spotlight metaphor for visual attention is a powerful one, but I believe it is based on a mistaken analogy between external perception and introspection. Normally our attention moves rapidly and smoothly from one level to another and we do not realise that at any instant we are attending at just one level. Only when the information at the different levels is made inconsistent, as in the fruitface, does it become obvious that the Gestalt for the whole cannot coexist with the Gestalts for its parts. Introspection is of little use for deciding what is in our minds at one brief instant because it does not allow us to decide between two possibilities. Either there are shape representations that lurk outside focal attention, or shape representations are generated or regenerated the moment we ask ourselves whether they are there. Our fundamental epistemological assumption that the existence of objects is independent of our awareness of them cannot be applied to the contents of our own minds.

An obvious objection to any theory which claims that people only see one shape at a time is that the shape of an object is determined by the shapes of its parts and their dispositions relative to the whole. This kind of recursive definition of a shape in terms of the shapes of its parts leads to a regress that only terminates at hypothetical "primitive" features. The fruitface figure is important because it suggests an alternative way out of the regress. The representations of the parts that are used in perceiving the shape of the whole may be different in kind from the representations used to perceive the shapes of the parts when we attend to them. Naturally, different shape representations must be able to influence one another. Having recognised an eye it should be easier to see the whole face, but this influence could be mediated by spatial working memory. Although only one Gestalt can be formed at a time, records of many previous Gestalts can be kept in working memory and used to influence the formation of the next Gestalt.

IV WHAT THE DEMONSTRATIONS SHOW

The demonstrations have been used as evidence for the following claims:

1. We integrate the information obtained in successive glances by storing records of the shapes that we identify and their relationships to a temporary scene-based frame of reference. We can use these stored records to generate new shape representations.

2. The process of recognising a shape (forming a Gestalt) involves imposing an object-based frame of reference and representing the size, position, and orientation of each part of the object relative to this frame.

3. The representation that an object receives when it is seen as a Gestalt and its shape is recognised is completely different from its representation when it is seen as a constituent of a larger Gestalt. Only one Gestalt can be formed at a time, but many separate records of previous Gestalts can be stored in spatial working memory.

V A MECHANISM FOR SPATIAL REPRESENTATIONS

There is not space here to discuss all the various kinds of mechanism that have been suggested for representing spatial structures. I shall simply describe one possibility which is designed to make use of parallel interactions between very large sets of features. This kind of computation seems to be a natural way of harnessing the computational power provided by a system like the brain in which a large number of richly interconnected units all compute in parallel (Anderson and Hinton, 1981). The mechanism is based on four related assumptions:

1. A perceptual feature must always be represented relative to some frame of reference because properties like the length, position, and orientation of a feature implicitly assume a reference frame.

2. At any moment during perception we use three different frames of reference -- retina-based, object-based, and scene-based -- so our perceptual apparatus has three different sets of units, each of which represents features relative to one of these frames of reference.

3. The meaning of features relative to one frame of reference in terms of features relative to another depends on the relationship between the two frames. So the way in which units in one set affect units in another set must be controlled by a representation of the spatial relationship between the frames of reference used by the two sets. A particular spatial relationship pairs each unit in one set with one unit in the other set, and allows activity in one of these units to cause activity in the other.

4. Different Gestalts correspond to alternative patterns of activity in the very same set of object-based units. So only one Gestalt can be formed at a time, though records of many previous Gestalts can be stored as activity in the scene-based units.

Fig. 4 incorporates these assumptions. Unlike many box diagrams in psychology, the separate boxes really are intended as separate collections of hardware units. Every unit continually recomputes its activity level as a function of the input it receives from other units. In the short term (i.e in about 100 msec), the whole system computes by settling into a state of activity that is temporarily stable. This kind of settling process is described in more detail in Hinton (1981b) where it is shown that the process of assigning an appropriate object-based frame of reference can be implemented by the three-way interaction between retina-based units, object-based units and the units for representing the spatial relationship between

the retina and the object. This kind of three way interaction is what the triangular symbols in Fig. 4 depict. After each settling, control processes (unspecified here) can reset the pattern of activity in any set of units, and thereby initiate a new process of settling. Not all the units in a set need be involved in the interactions with other sets. For example, the object-based units that are directly affected by retina-based units probably code fairly simple features, whereas the object-based units that directly affect the scene-based ones probably code complex conjunctions of the simpler features.

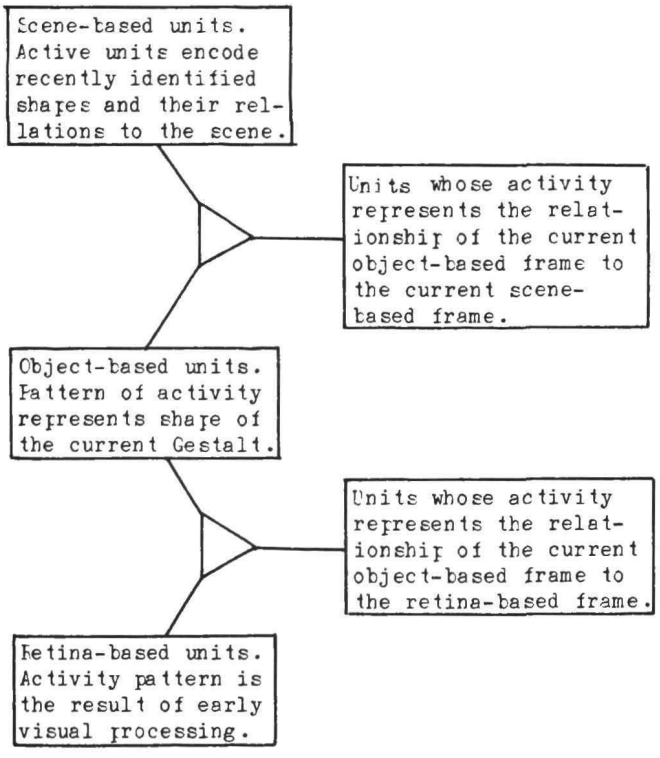


Figure 4. A parallel mechanism.

This kind of mechanism raises many interesting issues, some of which are discussed elsewhere (Hinton, 1981a). The following section focusses on what the scene-based features are like, and how they influence the the formation of a new Gestalt, i. e. how they affect the formation of temporarily stable pattern of activity in the object-based units.

Scene-based features

Once the general approach of implementing spatial working memory as activity in a set of scene-based units is accepted, quite a lot can be deduced about the nature of the units from their function. One important function of spatial working memory is to allow previously identified Gestalts to aid in the formation of related Gestalts. Having recognised an eye, the whole face should be easier to see, and vice versa. The kind of precisely located, atomistic features that would be needed for a picture-like representation would not be of much value in spatial working memory, because they would not explicitly represent the identities of objects, and so their effects could not be made to depend on these identities. It is more useful to make each active scene-based unit represent the existence of an object of a particular type with a particular relationship to the current scene, as the following examples show.

Suppose that as a result of previous perceptual analysis, activity in a scene-based unit, S_i, represents the existence of an eye with the relationship F_{is} to the scene. Suppose also that the system is now attempting to settle on an interpretation of a larger object (a face) with the relationship F_{fs} to the scene. F_{is} and F_{fs} determine F_{if}, the relationship of the eye to the face, and so they determine which object-based unit, C_i, should be activated to represent the eye as a constituent relative to the frame of reference of the whole face. This influence of the contents of working memory on perception can be implemented (see Fig. 4) by having an explicit representation of F_{fs} which governs the interaction between scene-based and object-based units and ensures that activity in S_i provides excitatory input to C_i.

Now consider what is required of spatial working memory if the face is seen first and attention is then focussed on one eye. The fact that this part had the role of an eye within the whole face should facilitate its interpretation as an eye when it becomes the focus of attention. This effect can be achieved if the Gestalt for the whole face activates scene-based units that represent the major constituents of the Gestalt as well as the whole. So the mapping from object-based to scene-based units operates simultaneously on units that represent the identity of the whole Gestalt and on units representing its major constituents.

VI CONCLUSION

Three demonstrations have been used to illustrate aspects of our internal representations of spatial structures. Particular attention has been given to the spatial working memory that allows people to integrate their perception over time. It has been argued that this memory contains compact records of the rich perceptual Gestalts that are formed when a person attends to an object. The interactions between spatial working memory and the apparatus in which Gestalts are formed allows previous Gestalts to influence (or entirely determine) the formation of the current Gestalt even though only one Gestalt can be present at a time. This view of the role of spatial working memory supports "early selection" theories in which focal attention is required to synthesize a shape, and only one shape can seen at a time. It also supports the view that different Gestalts correspond to alternative patterns of activity in a set of units that encode features relative to a frame of reference imposed on the object.

Finally, a few provisos. The demonstrations are well known but the interpretations of what they show are probably contentious, and the mechanism I suggest is speculative and underspecified. There has not been space to elaborate on many interesting issues like how the mechanism might account for the experimental data on mental rotation (Cooper and Shepard, 1973) or spatial working memory (Broadbent and Broadbent, 1981; Phillips and Christie, 1977). Nor has it been possible to discuss crucial

theoretical issues like the number of units that would be required by the mechanism, or the problems of encoding novel shapes in working memory.

ACKNOWLEDGEMENTS

I thank Steve Draper, Ed Hutchins, Tony Marcel, Don Norman, Dave Rumelhart, Tim Shallice, Joanne Sharp and Aaron Sloman for useful discussions. Many of the ideas presented here were developed while I was a Visiting Scholar at the Program in Cognitive Science at the University of California, San Diego, supported by a grant from the Sloan Foundation.

REFERENCES

Anderson J. A. & Hinton, G. E. Models of information processing in the brain. In G. E. Hinton & J.A. Anderson (Eds.) Parallel models of associative memory. Hillsdale, NJ: Erlbaum, 1981.

Broadbent, D. E. & Broadbent, M. H. P. Recency effects in visual memory. Quarterly Journal of Experimental Psychology, 1981, 33A, 1-15.

Cooper, L. A. & Shepard, R. N. Chronometric studies in the rotation of mental images. In W. G. Chase (Ed.), Visual information processing. New York: Academic Press, 1973.

Girgus, J. S. Gellman, I. H. & Hochberg, J. The effect of spatial order on piecemeal shape recognition: A developmental study. Perception and Psychophysics, 1980, 28, 133-138.

Hinton, G. E. Some demonstrations of the effects of structural descriptions in mental imagery. Cognitive Science, 1979, 3, 231-250.

Hinton, G. E. Shape representation in parallel systems. To appear in Proc. IJCAI-81 1981a.

Hinton, G. E. A parallel computation that assigns canonical object-based frames of reference. To appear in Proc. IJCAI-81. Vancouver, Canada, 1981b.

Hochberg, J. In the mind's eye. In R. N. Haber (Ed.) Contemporary theory and research in visual perception. New York: Holt, Rinehart and Winston, 1968

Marr, D. & Nishihara, H. K. Representation and recognition of the spatial organisation of three-dimensional shapes. Proc. Roy. Soc. Series B, 1978, 200, 269-294.

Palmer, S. E. Visual perception and world knowledge: Notes on a model of sensory cognitive interaction. In I. A. Norman & D. E. Rumelhart (Eds.), Explorations in cognition. San Francisco: Freeman, 1975.

Phillips, W. A. & Christie, D. F. M. Components of visual memory. Quarterly Journal of Experimental Psychology, 1977, 29, 117-134.

Rock, I. Anorthoscopic perception. Scientific American, March 1981, 244, 103-111.

Triesman, A. M. & Gelade, G. A feature-integration theory of attention. Cognitive Psychology, 1980, 12, 97-136.

Turvey, M. T. Contrasting orientations to the theory of visual information processing. Psychological Review, 1977, 84, 67-88.

COLOR PERCEPTION AND THE MEANINGS OF COLOR WORDS

Paul Kay
University of California, Berkeley

The relation between perception and meaning is hard to trace in any domain. I have been asked today to discuss the problems of identifying such connections in the domain of color. Color is an area in which our ignorance regarding the relation of perception and linguistic meaning is less than total; nonetheless you will not be surprised to learn that here, as elsewhere, there are more questions than answers. In the time available I will be able to do no more than sketch one view of the matter and so will probably present a clearer picture than is in fact warranted by current knowledge. In particular there will be little time to discuss the detailed empirical evidence that supports this view and no time to discuss alternative views.[1]

I will begin by describing in lamentably oversimplified terms certain structures in the human visual system that give rise to the sensation of color. We will then see how these aspects of visual physiology can help us understand independent findings regarding the meanings of words for color in the world's languages. In particular, starting from the simple arithmetic of differential firing rates of certain individual types of cells in the visual system, we are able to build a model of the perceptual categorization of color that explains a good deal both about cross-linguistically universal features in color naming and about dimensions of difference among the classifications of color found in the languages of the world. Finally we will see how this model organizes certain systematic observations that have been made regarding regularities in the temporal evolution of the color classification systems of the world's languages.

Aspects of the Neurophysiology of Color Perception

It is widely known that the retina contains three kinds of color receptors, i.e., cones. It is perhaps less generally known that at post-retinal but still peripheral levels of neural processing, information regarding dominant wavelength, or hue, is recoded from this three-channel system into a four-channel system, yielding the four fundamental hue sensations, blue, green, yellow, and red. In 1920 Ewald Hering (1968) postulated, on the basis of primarily introspective evidence, that such a system must exist. Hering noted further that subjectively there is no such thing as a mixture of green and red nor of blue and yellow--try to imagine what one could possibly mean by the locutions 'a reddish green' or 'a bluish yellow'. He therefore supposed that there must exist what he called two 'opponent processes' in the visual system, one red vs. green process and one yellow vs. blue process. At a given moment, each process has to be in exactly one of its named states; for example, at a certain time the red-green process might be in the red state and the yellow-blue process in the yellow state: this pairing of states would give rise to the sensation of orange. If one admits continuity to the model by allowing each of the four opponent states to operate at varying strengths, the relative strengths of the two states operative at a given moment will determine the precise shade that is subjectively experienced. In our example, the relative strengths of the red and yellow states will determine whether a reddish orange, a yellowish orange, or a relatively balanced or pure orange is experienced.

Despite the superiority of the Hering model over its competitors in explaining these and many other aspects of the subjective experience of color, it never gained general acceptance until Russell De Valois and his associates, as recently as the late 1960s, isolated the anatomical structures that accomplish the opponent-process function and monitored this function in the living organism. After a series of preliminary experiments, which established that the visual system of the macaque monkey is like man's in all relevant respects, De Valois and his co-workers inserted micro-electrodes into individual cells in the Lateral Geniculate Nucleii of live macaques and recorded the rates of firing of these cells while the animals' eyes were exposed to light of systematically varying wavelengths and intensities. It was found that LGN cells could be thus classified into six types, two of which were primarily sensitive to overall luminosity or brightness, and four of which were primarily sensitive to dominant wavelength or hue.

The latter four types constitute the opponent process system. Each of these opponent cells has a spontaneous or basal rate of firing (or about 6 spikes per second): this rate increases or decreases depending on the wavelength of stimulating light according to the four patterns shown in Figure 1.

Inspection of Figure 1 reveals that types of cell A and C form a pair in having the same crossover point between advanced and retarded rate of firing at about 605nm and also in having mirror image maxima and minima of firing rate at about 540nm and 640 nm. Cells of types A and C together constitute Hering's postulated red-green process: considering all cells of types A and C at once, we may take the sum of the absolute deviations from the basal rate of firing in the long wavelength region, that is above the crossover point, as signaling the strength of the red response. Similarly, the total absolute deviation from the basal firing rate below the crossover point represents the strength of the green response. In analogous fashion the type B and D cells together constitute the yellow-blue opponent process: the sum of absolute deviations above the crossover point is the total amount of yellow information or equivalently the strength of the yellow response, while below the crossover point the sum of absolute deviations represents the total blue response. Note that in a given stimulus condition, each opponent process must be in either one state or the other depending whether the wavelength of the stimulus is above or below the crossover point for that pair of types of cells.

At a given wavelength there are thus two families of possibilities. (1) If the visible wavelength is at one of the cross-over points, then one of the opponent systems is inert, e.g. at about 605nm the macaque's red-green system is quiescent; at this point all hue information is carried by the yellow-blue channel, which is in the yellow (longer wavelength) state; this is called the yellow unique hue point; all the organism sees at this wavelength is pure yellow. The blue, green, and red unique hue points are defined in the same way. (2) If the stimulus wavelength is not at a unique hue point, then exactly two of the four fundamental hue states are operative and the relative strengths of these two states determine the precise shade of perceived hue; for example in the region

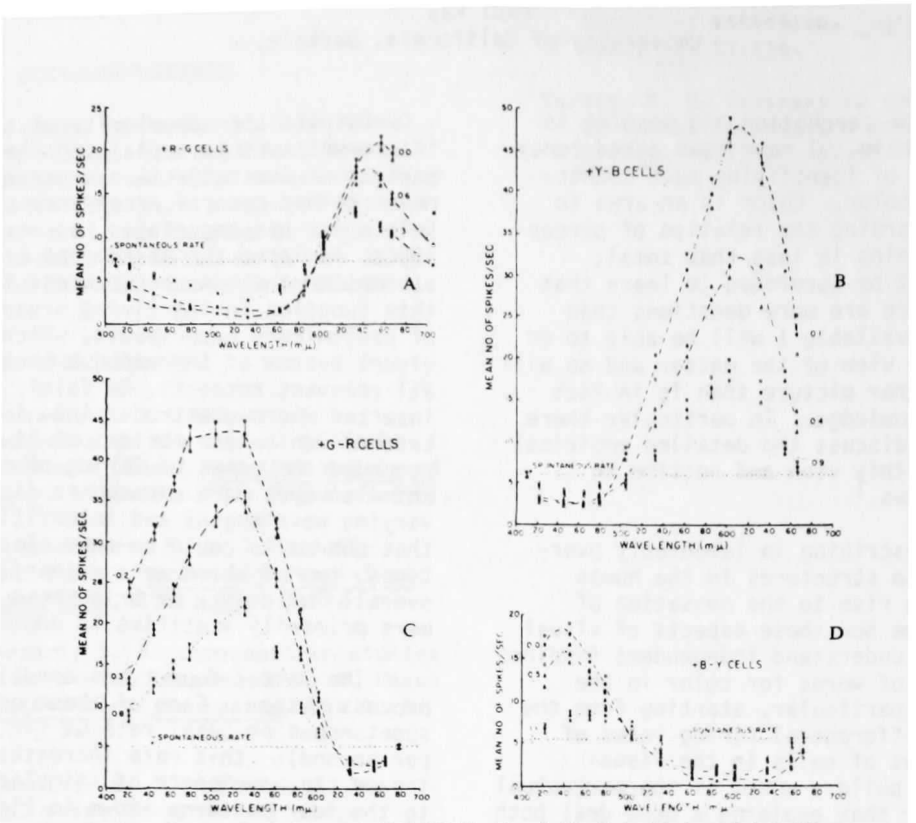


FIGURE 1. Source: De Valois et al. 1966: 972-73; Figs. 9-12). Note: Horizontal dotted lines indicate mean basal response rate for each cell type.

between the yellow and green unique hue points the green state of the red-green system and the yellow state of the yellow-blue system are operative; the relative strengths of these response states determine whether a yellowish green, a greenish yellow, or a perfectly balanced chartreuse or lime is perceived.

One may, in sum, model this system as having quantitative outputs in four channels, RED, YELLOW, GREEN, BLUE, where at a given instant there are non-zero outputs in either (1) a single channel or (2) two adjacent channels (considering red and blue as also adjacent). From psycho-physical data Wooten (1970) has estimated the curves for humans comparable to those of Figure 1. Using these curves (not shown here) it is a straightforward matter to calculate for each channel of fundamental hue response (RED, YELLOW, GREEN, BLUE) the proportion of total hue response in that channel for each wavelength of visible light. Curves representing these calculations are shown in Figure 2. Being proportions, these functions necessarily have ordinates varying from zero to unity across the spectrum. It is therefore natural to interpret them as fuzzy sets, which interpretation is reflected by the ordinate of Figure 2 being labelled "Degree of Membership".

So far we have talked only about types of neural cells, their rates of firing, and certain functions composed of the firing rates of different classes of cells, but we have said nothing about the meanings of any words in natural languages. We are now prepared to make the initial connection: the curves labelled BLUE, GREEN, YELLOW, and RED in Figure 2 represent at one and the same time (1) the outputs of the fundamental hue-response categories as defined in terms of proportional output in the individual hue channel of the opponent process system, and also (2) the meanings of the ordinary English words blue, green, yellow, and red, along with their exact translations into many languages. Similarly, non-opponent fundamental response channels BLACK and WHITE, corresponding to the English words black and white (and their translations in many other languages), are defined by the two classes of brightness-sensitive cells discovered by De Valois and his associates, and these categories also may be modeled as fuzzy sets.

The Semantics of Color Words

We have seen that six English color words (and their translations into other languages that have exact translations of these words) can be given

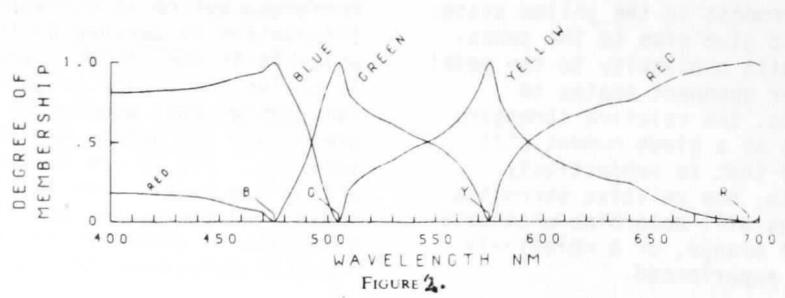


FIGURE 2.

neurophysiological definitions. For these six semantic categories, we have achieved a considerable rapprochement of semantics and perception. What can we say now about the perceptual basis of other color words in English and in other languages? In this investigation we will restrict our attention to what have come to be called 'basic' color words. In any language the basic color words form a natural set, and it is the comparison of the sets of basic color words across languages that has been found most fruitful in the cross-linguistic investigation of this semantic domain. In every language there is a small set of semantically simple words such that any color can be named with a member from this set. Members of this set are called the basic color words or basic color terms of the language. Several languages are known in which there are just two basic color terms. English has eleven; in addition to the six already discussed, which name the fundamental neural response categories, there are also brown, purple, pink, orange and grey. For many speakers of Russian, there are twelve basic color terms; Russian has a basic color term specifically for light blue, goluboy, along with the term for darker blue, sinyiy.

For a long time it was believed by linguists and anthropologists that there were no constraints on the way the basic color terms of a language might divide the perceptual domain of color and hence no tendency for color words to be translatable across unrelated languages. Another way this idea was put was the claim that perception has no influence over color-naming in a language beyond setting the bounds of the visible spectrum. Thus in what was probably the most widely accepted linguistics textbook of the 1950s, H. A. Gleason said, "There is a continuous gradation of color from one end of the spectrum to the other. Yet an American describing it will list the hues as red, orange, yellow, green, blue, purple--or something of the kind. There is nothing inherent either in the spectrum or the human perception of it which would compel its division in this way" (1961:4). We now know that this is wrong and that all the basic color terms in all languages are based on the six fundamental response categories: the four of the opponent (i.e. hue) system and the two non-opponent (i.e. brightness) categories.

We have already noted that each of these six categories has a structure that invites its interpretation as a fuzzy set. There is strong additional motivation for the fuzzy set interpretation, namely that all the other basic color categories, either in English or in any of the other languages that have been investigated, may be defined in terms of simple Boolean functions of these fuzzy sets. For example, in many languages of the world, including the majority of Native American languages, there is a single word that is used wherever an English speaker would use either the word green or the word blue. There is considerable experimental evidence indicating that this widespread basic color category (let us call it 'grue') is in fact the fuzzy union of the fundamental neural response categories GREEN and BLUE. For example, it has been found in a large number of languages that subjects asked to pick out from an array of color stimuli the best example of their category grue will not select something intermediate between green and blue such as we might call turquoise or aqua; rather, they will select either a focal green or a focal blue. Since the union of two fuzzy sets is defined as the maximum of the individual characteristic functions, this pre-theoretically surprising, but empirically robust, finding is predicted by the definition of the category 'grue' as the fuzzy union of GREEN and BLUE.

Berlin and Kay (1969) surveyed the basic color lexicons of ninety-eight languages and reported strong constraints on the semantics of basic color term systems. They also postulated a narrowly constrained evolutionary sequence through which basic color lexicons must pass as they add terms over time. That sequence, as reformulated by Kay and McDaniel (1978) about a decade later on the basis of a great deal of work by many investigators in the interim, is summarized in Figure 3.

A language with only two basic color terms has one which is the fuzzy union, 'WHITE or RED or YELLOW', and one which is the fuzzy union, 'BLACK or GREEN or BLUE'; these are conveniently glossed as 'light-warm' and 'dark-cool', respectively. When a language adds a third term, it does so by splitting the 'light-warm' term into a 'WHITE' term and a 'RED or YELLOW' (i.e. 'warm') term. At the next stage of development, either the 'dark-cool' term splits into 'BLACK' and 'GREEN or BLUE', that is 'cool', or the warm term splits into 'RED' and 'YELLOW' terms (see Stages IIIa and IIIb in Figure 3). At Stage IV, whatever possibility didn't occur at Stage III now occurs, so the language now has basic terms for the fuzzy categories 'WHITE', 'BLACK', 'RED', 'YELLOW', and 'GREEN or BLUE' (i.e. 'grue'). At Stage V, the 'grue' category is dissolved into its fundamental neural response components 'GREEN' and 'BLUE', and there is now one basic color term for each fundamental neural response category.

Up to here in the sequence, we have been considering two types of basic color categories, those that consist in unions of fundamental neural response categories and those that consist in the fundamental neural response categories themselves. Beyond evolutionary Stage V, basic color categories of a new kind are formed on the basis of the intersections of the fundamental categories. More precisely each of

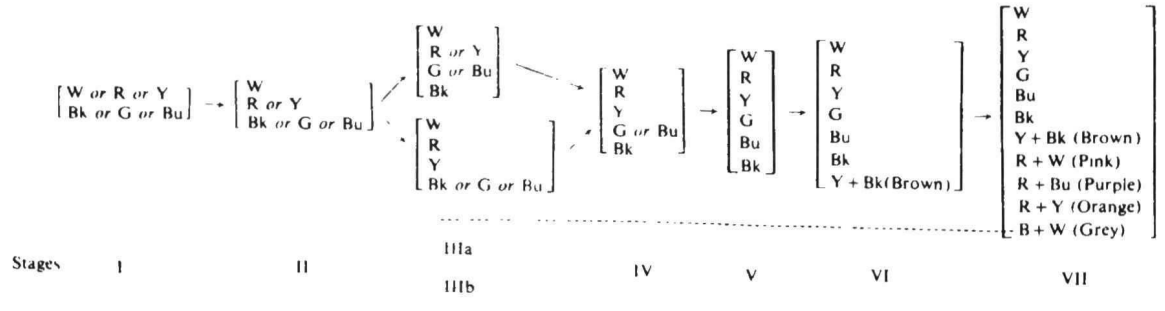


FIGURE 3

these later combinations of the fundamental categories consists in twice the fuzzy intersection of its constituent categories. For example, the fuzzy set orange is twice the intersection (minimum) of the fuzzy sets RED and YELLOW.[2]

It is not possible in the time available to discuss the empirical motivation for the formulation of these intersectional categories in terms of precisely twice the intersection of the constituent fuzzy categories (see Kay and McDaniel 1978:631-635; Mervis and Roth 1981). But the main points of the story so far should now be clear. Empirical semantic researches have revealed that, so far as we can tell at present, all the basic color categories of the languages of the world are based on the six fundamental neural response categories, whose structures are determined by the firing patterns of LGN (and other) cells in the visual pathway. Languages with fewer than six basic color terms have terms that encode categories composed of fuzzy unions of the fundamental categories. Languages that encode more basic categories than the six perceptually fundamental ones, encode categories based on the fuzzy intersections of the fundamental ones.

Furthermore, there appear to be quite narrow constraints on which of the logically possible Boolean combinations of the six fundamental response categories actually occur in the world's languages. For example, of the fifty-seven possible categories that might be formed by taking fuzzy unions of the six neurologically fundamental categories, only the four we have discussed ('light-warm', 'warm', 'dark-cool', and 'cool'--i.e. 'grue') occur in actual languages. Little is known by way of explanation of this fact, though it is perhaps worth recalling that Hering designated the colors white, red and yellow collectively as inherently arousing, and the colors black, green and blue as inherently non-arousing. Even more striking as an empirical generalization crying for theoretical explanation is the evolutionary sequence depicted in Figure 3. Why should the color lexicons of the world sort into just the handful of types permitted by this sequence and, above all, why should the temporal evolution of color terminology systems follow this particular, narrowly restricted course? Answers to these questions, as they are found, will deepen our understanding of the relation of perception and linguistic meaning in the domain of color.

Notes

1. This talk is, in effect, a highly compressed summary of Kay and McDaniel (1978), and the hearer or reader interested in pursuing the subject should consult that paper and the references cited there. McDaniel (1972) was the first to propose a perceptual explanation for the Berlin and Kay (1969) findings regarding semantic universals in terms of the opponent process model of color vision. Figures 1, 2, and 3 accompanying this text are respectively Figures 4, 6, and 13 of Kay and McDaniel.

2. In Figure 3, the '+' sign denotes the binary operation 'twice the fuzzy intersection'.

References

Berlin, B., and Paul Kay. 1969. Basic color terms: their universality and evolution. Berkeley & Los Angeles: University of California Press.

De Valois, R. L.; I. Abramov; and G. H. Jacobs. 1966. Analysis of response patterns of LGN cells. Journal of the Optical Society of America 56.966-77.

_____, and G. H. Jacobs. 1968. Primate color vision. Science 162.533-40.

Gleason, H. A. 1961. An introduction to descriptive linguistics. New York: Holt, Rinehart and Winston.

Hering, Ewald. 1920. Grundzuge der Lehre vom Lichtsinn. Berlin: Springer. [English version: Outlines of a theory of the light sense. Translated by L. M. Hurvich & D. Jameson. Cambridge, MA: Harvard University Press, 1964.]

Kay, Paul, and Chad K. McDaniel. 1978. The linguistic significance of the meanings of basic color terms. Language 54:610-646.

Mervis, C. B., and E. M. Roth. 1981. The internal structure of basic and non-basic color categories. Language 57:383-405.

Wooten, B. R. 1970. The effects of simultaneous and successive chromatic constraint on spectral hue. Doctoral dissertation, Brown University.

Structure and Function in the early processing of visual information

Shimon Ullman

The Artificical Intelligence Laboratory
Massachusetts Institute of Technology

1. Introduction

A central notion in contemporary cognitive science is that mental processes involve computations defined over internal representations. This general view suggests a distinction between the study of the representation and computations performed by our cognitive systems on the one hand, and the physical brain mechanisms supporting these computations on the other. The two studies proceed along different paths, and neither is completely reducible to the other. It is the hope of cognitive science, however, that the studies of function and mechanism can complement each other, and that theories can be developed for various cognitive subsystems that will describe and explain their computational aspects, their underlying mechanisms, and the interactions between the two.

In this paper I shall describe some attempts to combine the study of brain mechanisms with computational considerations in the first stages of visual information processing. This work combines the contributions of many individuals, most notably the late David Marr, and a group of people who were fortunate to work with him, primarily at M.I.T's Artificial Intelligence Laboratory and Psychology Department.

2. Representing intensity changes in images

The first computational problem that arises in the early processing of visual information is the initial organization and representation of the input registered by the eyes. At the photoreceptors level, the input to the visual system consists of over 250 million light intensity measurements (registered by over 120 million cones and rods in each eye). This is an unwieldy large and unstructured set of measurements. We can therefore expect the visual system to construct a more economical representation of the input, that will make explicit the relevant information for later processing stages.

A reasonable candidate for the task is a representation that can be roughly described as an edge representation of the image. The idea is to make explicit the locations in the image where light intensity changes sharply from one level to another. The motivations for this type of a representation are (i) it will achieve a more concise description of the image than the original array of intensity values , and (ii) sharp changes in light intensity values usually have a physical significance. They are often associated, for example, with object boundaries, markings on objects' surfaces, and so forth An edge representation is therefore useful in making the transition from the domain of light intensities in the image to analyzing the physical structure of the visible environment. One general observation often raised in support of the edge representation approach is that many objects are recognizable from a sketch of their edges and contours alone, although in terms of the underlying light intensity distributions, the sketch and the original image are markedly different.

The representation of localized intensity changes is not the only approach that has been proposed for the first stages of analyzing visual information. One popular alternative is the Fourier analysis approach that received wide attention in the psychophysical literature following Campbell and Robson's [1968] discovery of spatial frequency tuned channels in the visual system. The approach presented here is in a sense a combination of the frequency channels and the edge detection approaches, but it is concerned primarily with the detection of intensity changes.

A large variety of techniques have been proposed in the past (primarily within the engineering field of image processing) for the detection of intensity changes in images. A major problem that has been discovered in the course of developing these techniques, is that significant intensity changes in an image can occur at a variety of scales. Some changes are gradual and smooth: they can also be described in frequency domain terminology as low frequency changes. Others are high frequency and sharply localized changes. To capture all of the significant intensity changes, it is possible to examine the image at a number of different resolutions, or scales. A low resolution "copy" will serve for capturing the gradual, gross changes, a high resolution "copy" for the fine details. Figure 1 shows an example of what it means for the same image to be examined at three different resolutions. The resolution decreases from 1a to 1c. It can be seen that in the lower resolution copies fine details are progressively blurred. The low resolution copy can be obtained by a process called gaussian filtering (and this filtering is in a sense optimal, see Marr & Hildreth 1980). This simply means that at every point a local average is taken of the intensity values, using a gaussian weighting function. The resolution of the resulting copy is controlled by the size of the gaussian. A larger guassian averages the intensity values over a wider neighborhood, and hence is less sensitive to fine details. The gaussian smoothing is also called in mathematical terms the convolution of the image with a gaussian filter, denoted by $G*I$ (where I is the image, G is the gaussian smoothing function).

As a result of the first operation we have a number of "copies" of the original image, at a number of different resolutions, as determined by the sizes of the gaussian filters (figure 2). The next step is to isolate the sharp intensity changes in each copy. We shall consider this problem first in the context of one-dimensional signals. In this case, the image I is a function of a single variable, denoted by x. A sharp change in the signal $I(x)$ can be defined as a peak in its first derivative, since the derivative, by definition, measures the signal's slope. From elementary calculus, peaks in the first derivative can also be located by zero-crossings of the second derivative (i.e. places where the second derivative changes sign). Mathematically the two criteria are equivalent, but the second characterization has certain advantages when two-dimensional signals are concerned [Marr & Hildreth 1980].

In summary, the localization of sharp changes is obtained by performing:

$$\frac{d^2}{dx^2}(G*I) \qquad (1)$$

The zero crossings in the output will indicate the locations of sharp intensity changes in the image at the scale determined by the gaussian.

This means that the image I is first passed through a gaussian smoothing function G, and then a second derivative of the result is taken. The two operations of scaling and differentiation be combined in a convenient manner. The combination is based on a mathematical identity that states that the order of differentiation and convolution can be changed without affecting the result. In mathematical notation:

$$\frac{d^2}{dx^2}(G*I) = (\frac{d^2}{dx^2}G)*I \qquad (2)$$

The implication is that the two operations can be collapsed into a single one: simply filter the image not through a gaussian function, but through $\frac{d^2}{dx^2}G$ (the second derivative of a gaussian). This function is shown in figure 3a (3b shows its fourier transform). The analogue in two dimensions would be a similar but circularly symmetric function which has the appearance of a "mexican hat". Mathematically, in the two-dimensional case the filter is $\nabla^2 G$, where ∇^2 is the laplacian and G a two-dimensional gaussian function.

The scheme is now straightforward: the representation of intensity changes is obtained from the zero-crossing in the result of passing the image through filters that have the shape of $\nabla^2 G$.

Those who have some familiarity with the physiology of the visual system would readily recognize the shape of these filters as corresponding to the shape of retinal ganglion receptive fields. In other words, the retinal structure can be viewed as approximating the convolution of the image with the $\nabla^2 G$ filters. (For more detail see Marr & Hildreth 1980, Marr & Ullman 1981).

Figure 4 shows examples of images following this retinal operation, and the resulting zero-crossings representations (generated by Ellen Hildreth). The first row shows two images prior to the filtering stage. The second row shows the images filtered through the retinal operation. It gives some idea of the form of the image as it travels up the optic nerve from the eye, via an intermediate station called the LGN, to the visual cortex in area 17 of the brain. The third row illustrates the resulting zero-crossing representations. Figure 5 shows an image (of a sculpture by Henry Moore) and its zero-crossing representation at three different resolutions.

Before turning to the physiological aspects of the zero-crossing representations, it will be of interest to note that zero-crossings in bandpass filters are known to be, in a sense, "rich in information". B. Logan of the Bell Laboratories has shown that a one-dimensional signal with a bandwidth of less than one octave can be completely reconstructed (up to an overall multiplicative constant) from its zero-crossings alone, provided that some simple conditions are met [Logan 1977]. It is not clear, however, whether the theorem can be extended to two dimensions, and under what conditions the one-octave restriction can be relaxed (this problem arises since the filters in the human visual system are probably more than an octave wide). If appropriate extensions along these lines can be made, it would imply that the zero-crossings provide not only a convenient representation that captures the significant aspects of the image, but also a complete one. That is, no essential information is lost by discarding the image and analyzing the zero-crossing representation alone. (See Marr, Poggio & Ullman 1979, for further discussion of this issue.)

3. The biological detection of zero-crossings

The analysis so far leads to the general suggestion that following the retinal operation the next step is to locate and represent a map of the zero-crossings in the output. If this suggestion is correct, then a main function of the primary visual cortex should be the construction of the zero-crossings representation. I shall next turn to consider briefly how zero-crossings may be detected by the mechanisms of the visual cortex.

The fibers of the optic nerve coming from the eye to the brain carry the image filtered through the $\nabla^2 G$ receptive fields (this is, of course, a computational idealization). This neural image is in fact carried by units of two complementary types, called on-center and off-center units. The off-center units are simply "inverted mexican hats" with negative center and positive surround. Let us now consider the retinal output in the vicinity of an edge. Figure 6a depicts a step edge, and 6b is the result of passing 6a through retinal-like receptive fields. This output contains both negative and positive values. In contrast, the optic nerve carries no negative values; the positive part of the signal is carried by the on-center units, and the negative part by the off-center ones. This means that within the system the zero-crossing itself is always flanked by two peaks of activity: of on-center cells on one side, and off-center cells on the other. The detection of a zero-crossing can easily be accomplished, therefore, by a simple combination of the on- and off-center units. When two adjacent units, one off-center, the other on-center, are active simultaneously, they indicate the existence of a zero-crossing running midway between them. Note that a point of zero value is detected in this scheme by detecting peaks of activity rather than zero activity.

The basic zero-crossing detector is shown in figure 7a. It is composed of the two sub-units (on- and off-center) combined with an "and" operation. This means that the two units are required to be active simultaneously to produce a response. The unit can be made oriented by combining a number of such detectors lying in a row (figure 7b). Such an oriented unit will exhibit many of the properties of cortical simple cells ("edge detectors") originally discovered by Hubel & Wiesel in the visual cortex of the cat [1962] and monkey [1968]. It will still lack, however, one fundamental property: cells in the visual cortex are also often selective for direction of motion. They respond well when their preferred stimulus moves in one direction, but little or not at all when it moves in the opposite direction.

4. Adding directional selectivity

With the addition of one subunit it is possible to make the basic zero-crossing detector directionally selective, and use it for the measure-

ment of visual motion. To see how, consider again the zero-crossing associated with an intensity edge (figure 6b). At the zero-crossing itself the current value is, of course, zero. It can be readily seen from the figure that if the profile now moves to the right, the value at this point will be increasing. If it move to the left, the value will be decreasing. By simply inspecting the sign of the temporal change it therefore becomes possible to determine the direction of motion. It is not difficult to establish that it is further possible to measure the speed of motion in the direction of the unit by comparing the slope of the zero-crossing and the rate of temporal change. The extra sub-unit should respond therefore to temporal changes. Ideally, it should behave like the time derivative of the signal, i.e. $\frac{d}{dt}(\nabla^2 G)$.

As it turns out, the population of retinal cells contain a natural candidate for this task. These are the so-called Y-type cells, originally discovered by Enroth-Cugell & Robson [1966]. This is a relatively small sub-population of cells that are known to be "transient". That is, they respond to a steady stimulus by a short and brisk response when the stimulus is turned on or off. The other major population of retinal cells are the sustained, X-type cells. Such a cell responds to a stationary stimulus with a sustained response that usually continues as long as the stimulus is present within its receptive field.

Our schematic model of the simplest directionally selective units is therefore constructed from three types of sub-units. As before, it has a row of on-center cells, and a row of off-center cells, both of the sustained type. In addition, it has an input from at least one transient Y-type unit (figure 7c). A more detailed discussion of this general scheme can be found in [Marr & Ullman 1981].

This general scheme for zero-crossing and motion detection was driven primarily by computational considerations. Physiologically, although Y-type units were often described as transient, it was not clear whether they can also be described as at least approximating the required time derivative operation. We therefore compared the response required by the computational scheme with physiological response (taken from Rodieck & Stone, 1965, Dreher & Sanderson 1973; see Marr & Ullman 1981 for details). Some comparisons are shown in figure 8 (for X cells) and 9 (for Y cells). The top row in figure 8 is the convolution of various profiles (edge, thin bar, wide bar) with $\nabla^2 G$. On-center cells are expected to carry the positive part of these profiles, and off-center the negative part. In the next two rows the positive part of the signal is compared with recordings from on-center cells, and in the last rows the negative part is compared with recordings from off-center cells. Similar comparisons are shown in figure 9 between the computational model, based on $\frac{d}{dt}(G*I)$, and physiological recordings. It can be seen that even in the cases where the profiles are rather complicated, the general agreement is good.

Finally in this section, figure 10 shows an example of applying the motion detection scheme described above to a moving random texture. Figures 10a and b show a pair of random dot patterns. A central square in 10a is shifted in 10b slightly to the right, while the backgrounds of the two figures are uncorrelated. When these figures are presented to human observers in a rapid alternation, the central square is immediately perceived to move back and forth against a background of uncorrelated motion. Figure 10c shows the zero-crossings representation of 10a. Figure 10d is the result of the motion analysis of the zero-crossing (the light dots indicate the direction of motion of the zero-crossings). In figure 10e the light dots where removed from the area where coherent motion (to the right) was found. The motion assignment was correct, with the exception of a few isolated points, and as a result the moving square was detected.

I have sketched above some aspects of an evolving theory of early visual information processing. The main goal has been not to present a comprehensive review of the theory, but to illustrate an attempt aimed at combining the study of structure and function in the early stages of visual perception. Major parts of the theory were consequently left out of the discussion, most notably, the use of the early representations in stereo vision [Marr & Poggio 1979, Grimson 1981].

Finally, I would like to end with two brief cautionary notes. The first has to do with the specific problem of analyzing image contours. Even if the zero-crossing analysis is along the right track, it provides only the first stages in the analysis of edges and image contours. Figure 11 illustrates examples of contours that are easily perceived but cannot be captured by any simple intensity-based analysis of the image. In figure 11a all the lines lie along the 45 deg. diagonals. The horizontal and vertical boundaries which are apparent in the image are produced not by abrupt intensity changes, but by certain grouping processes. Figure 11b is an example of so called "cognitive contours". They do not exist in the image, and cannot be detected by simple intensity-based operations. These example serves to illustrate that even a seemingly simple and elementary task such as the detection of image contours, requires in fact complex processing that is still far from being completely understood.

The second and more general comment has to do with the integration of theories of function and structure in more complex systems. The examples I have outlined come from a system that is relatively simple and easy to explore. Its anatomical structure is orderly, the input to the system is relatively easy to control and manipulate experimentally, and much is known about its physiology. Even under these favorable conditions, the integration of structure and function proves to be exceedingly difficult. What is the hope, then, for achieving comprehensive theories of structure and function for higher, more complicated, cognitive systems?

The task is certainly formidable, but it is probably worthy of exploration at least in certain instances, since it appears unlikely that the structure of complex systems can be understood without some guidelines supplied by computational theories. It has to be admitted, however, that given the difficulties of the task it is unclear whether coherent and detailed theories combining structure and function can be achieved at present beyond the simplest cognitive systems.

Acknowledgment: I wish to thank E. Hildreth and K. Stevens for their invaluable help.

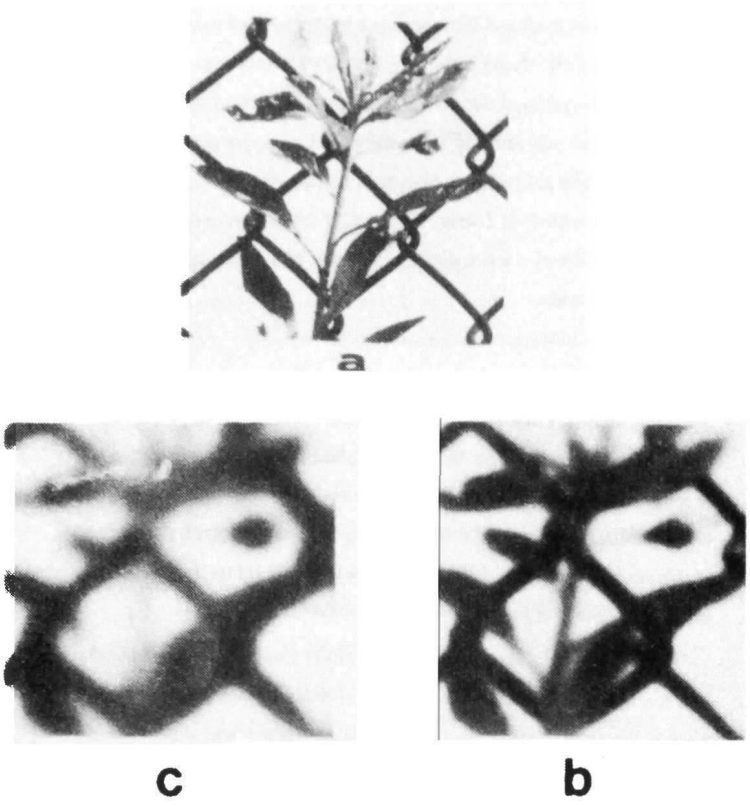


Fig. 1

The same image at three different resolutions.

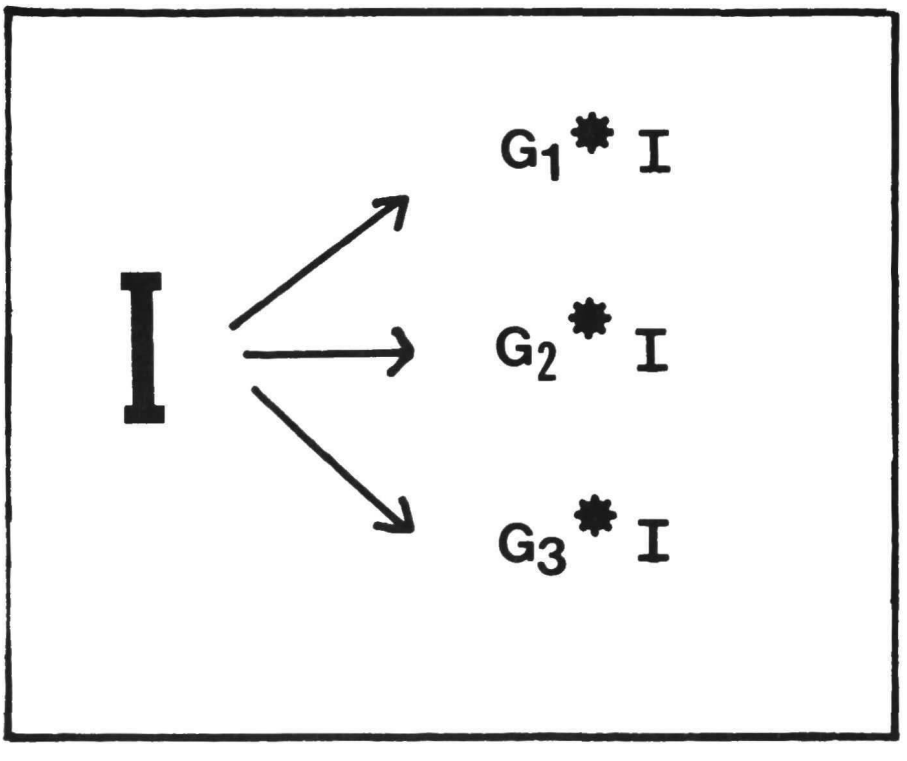


Figure 2. Different resolution copies of the original image are obtained by convolving the image with gaussian filters of different sizes.

Campbell, F.W. & Robson, J.G. Application of Fourier analysis to the visibility of gratings. *J. Physiol. (London) 197,* 551–556.

Dreher, B. & Sanderson, K.J. 1973 Receptive field analysis: responses to moving visual contours by single lateral geniculate neurons in the cat. *J. Physiol., Lond. 234,* 95–118.

Enroth-Cugell, C. & Robson, J. D. 1966 The contrast sensitivity of retinal ganglion cells of the cat. *J. Physiol. (Lond.) 187,* 517-522.

Grimson, W.E.L. 1981 A computer implementation of a theory of human stereo vision. *Phil. Trans. Roy. Soc., B. 292 (1058),* 217–253.

Hubel, D.H. & Wiesel, T.N. 1962 Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *J. Physiol. London, 160,* 106–154.

Hubel, D.H. & Wiesel, T.N. 1968 Receptive fields and functional architecture of monkey striate cortex. *J. Physiol. London, 195,* 215–243.

Logan, B.F. 1977 Information in the zero–crossings of bandpass signals. *Bell Sys. Tech. J., 56,* 487–510.

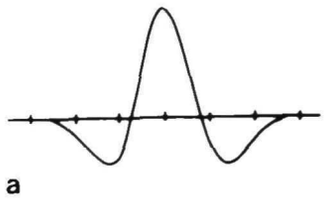
Marr, D. & Poggio, T. 1979 A computational theory of human stereo vision. *Proc. Roy. Soc. Lond. B 204,* 301–328.

Marr, D. Poggio, T. & Ullman, S. 1979 Bandpass channels, zero–crossings, and early visual information processing. *J. Opt. Soc. Am., 69(6),* 914–916.

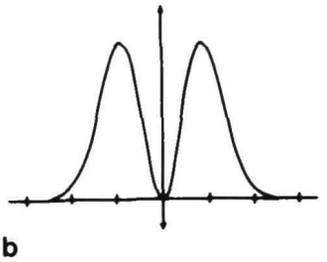
Marr, D. & Hildreth, E. 1980 Theory of edge detection. *Proc. R. Soc. Lond. B,* 187–217.

Marr, D. & Ullman, S. 1981 Directional selectivity and its use in early visual processing. *Proc. Roy. Soc. Lond. B, 211* 151–180.

Rodieck, R. W. & Stone, J. 1965 Analysis of receptive fields of cat retinal ganglion cells. *J. Neurophysiol. 28,* 833-849.



a



b

Figure 3. *a*. The shape of $\frac{d^2}{dx^2}G$. *b*. Its Fourier transform.

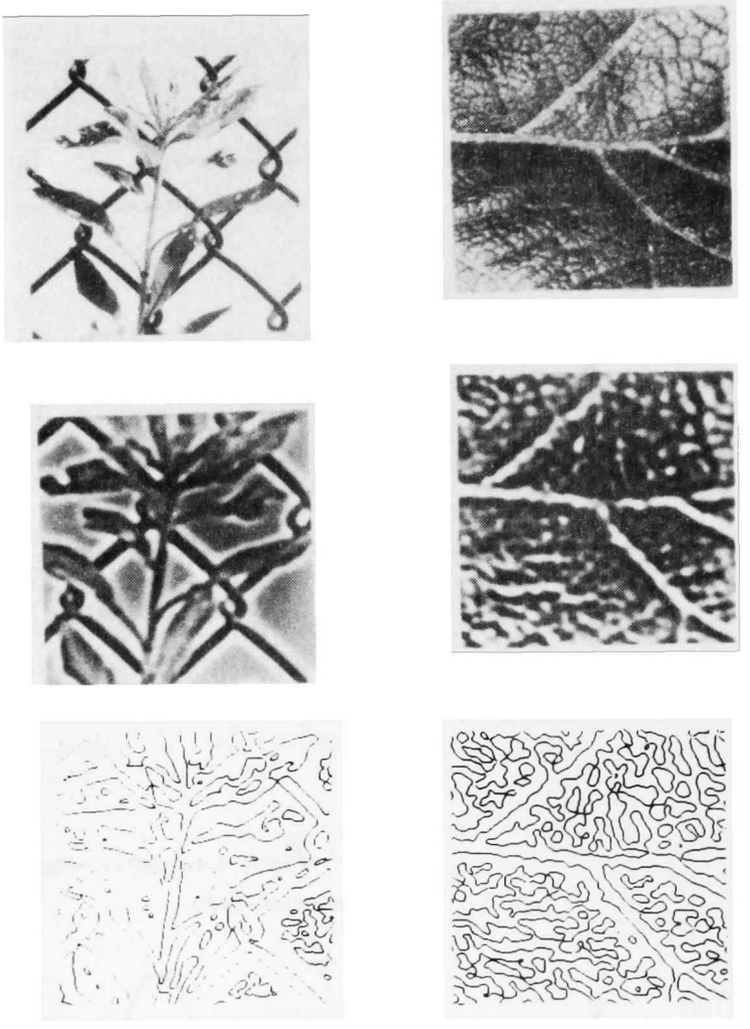


Figure 4. Examples of zero-crossing representations. First row: the original images. Second row: the images following the convolution with $\nabla^2 G$. Third rwo: the resulting zero-crossings representations.

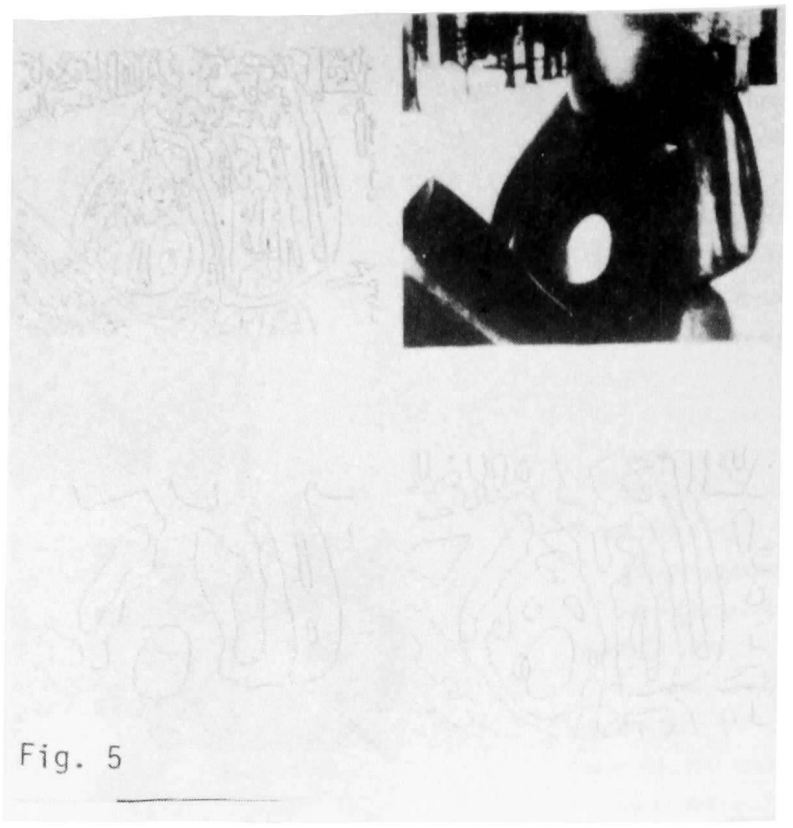


Fig. 5

Zero-crossing representations of the same image at three different resolutions.

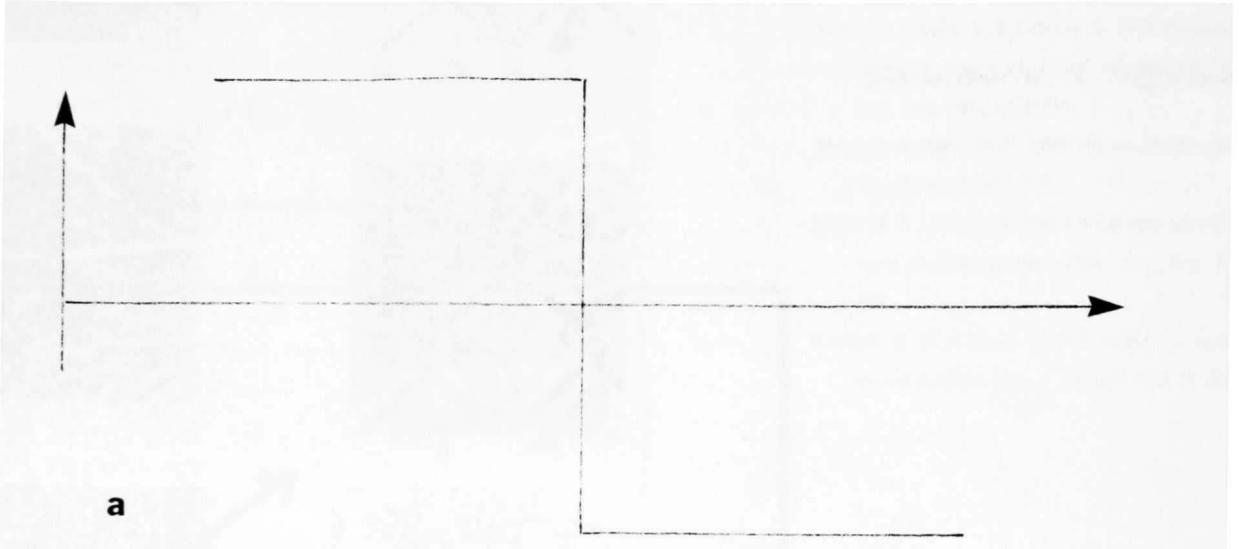
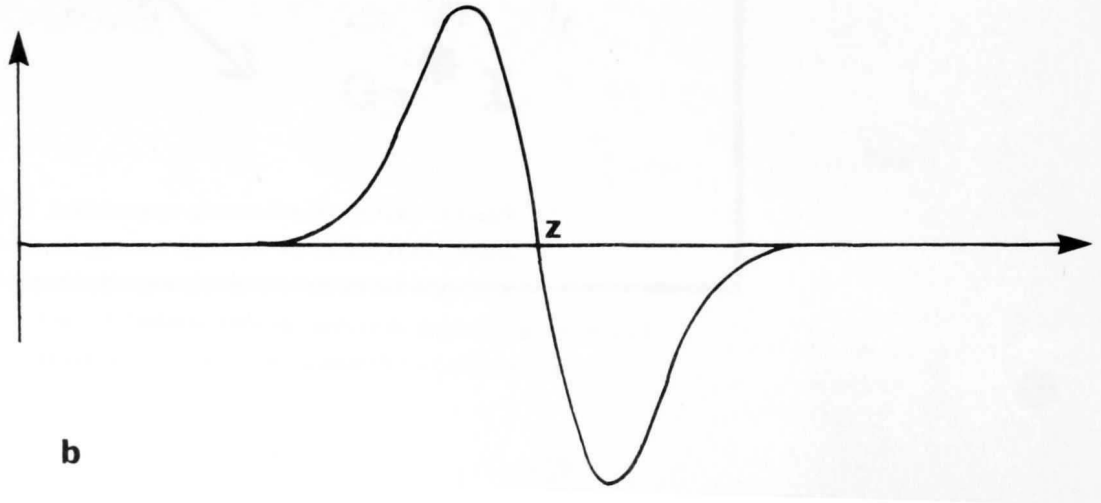


Figure 6. A step edge before (a) and after the convolutiion with the retinal operator (b).



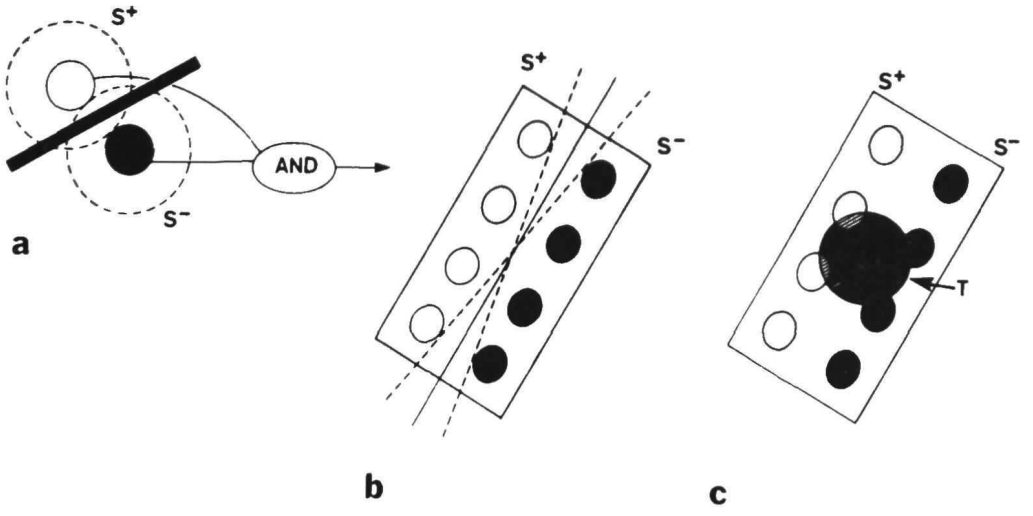


Figure 7 A schematic diagram of the basic zero-crossing detector. *a*. On-center and off-center units are *AND*ed together. *b*. A row of such subunits makes the detector orientation-specific. *c*. With the addition of a time derivative subunit the detector becomes directionally selective.

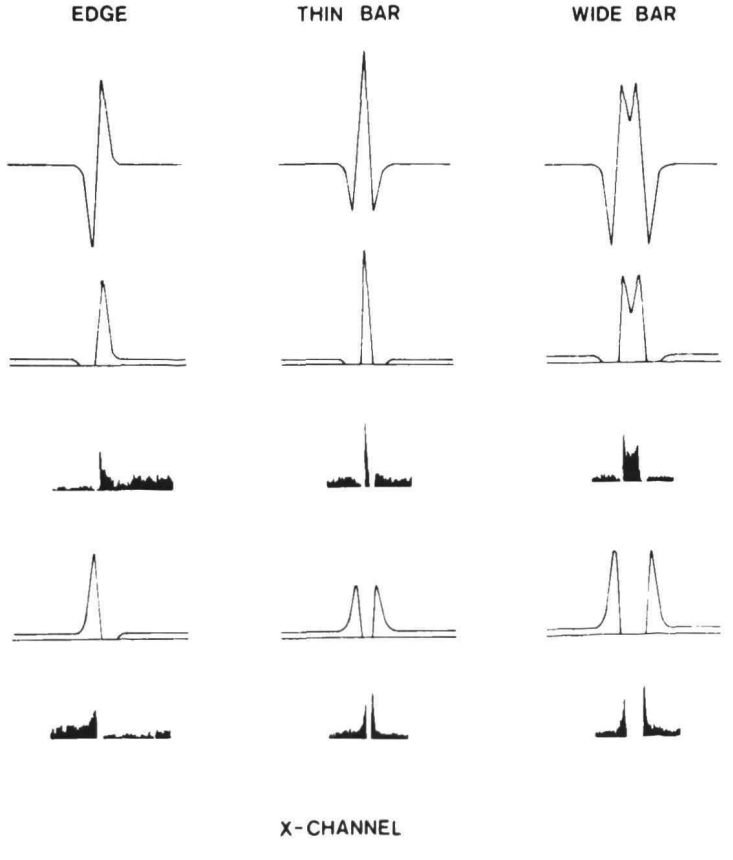


Figure 8. A comparison between the computational model and physiological recordings, for on- and off-center X-type units. For details see text

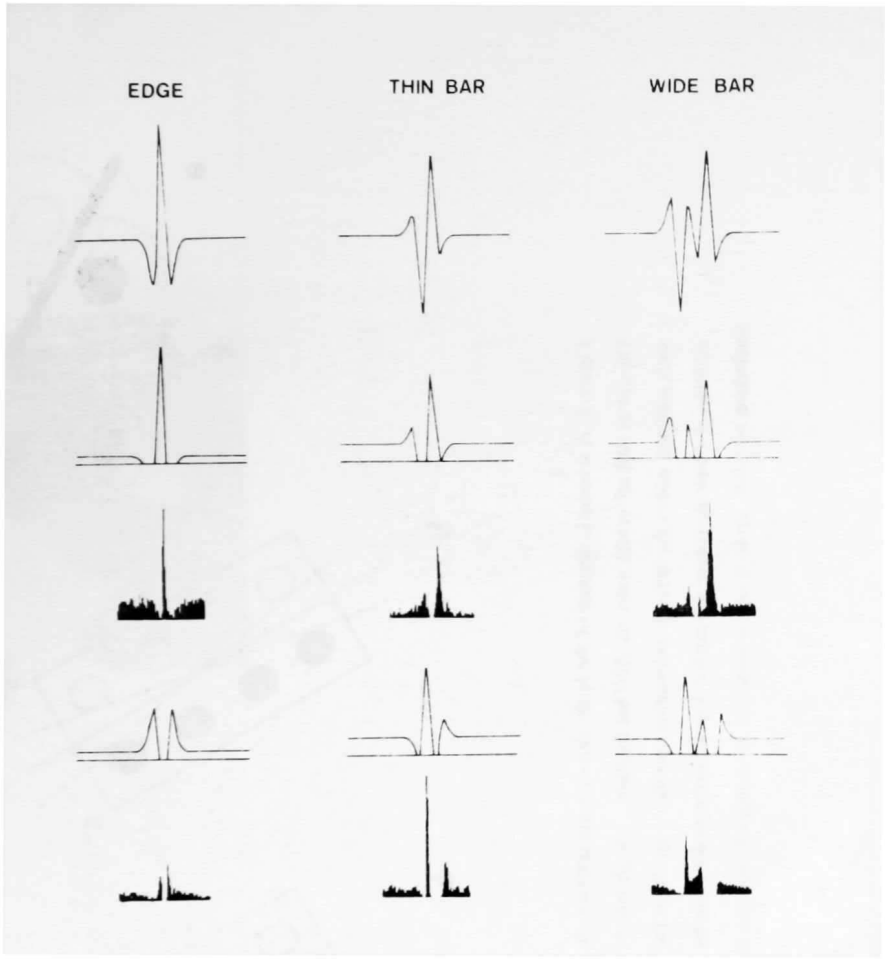


Figure 9. A comparison between the computational model and physiological recordings, for on- and off-center Y-type units. For details see text.

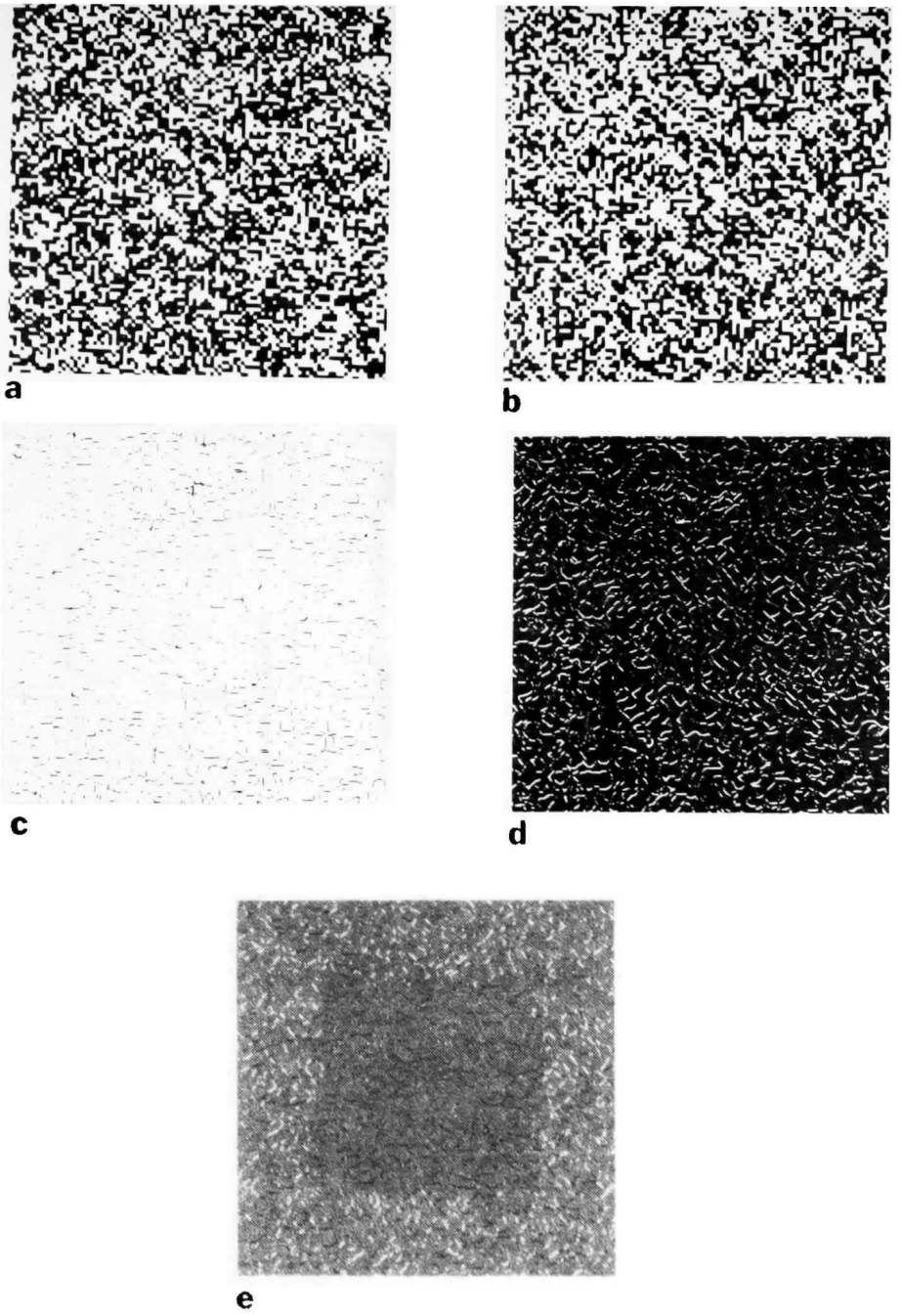
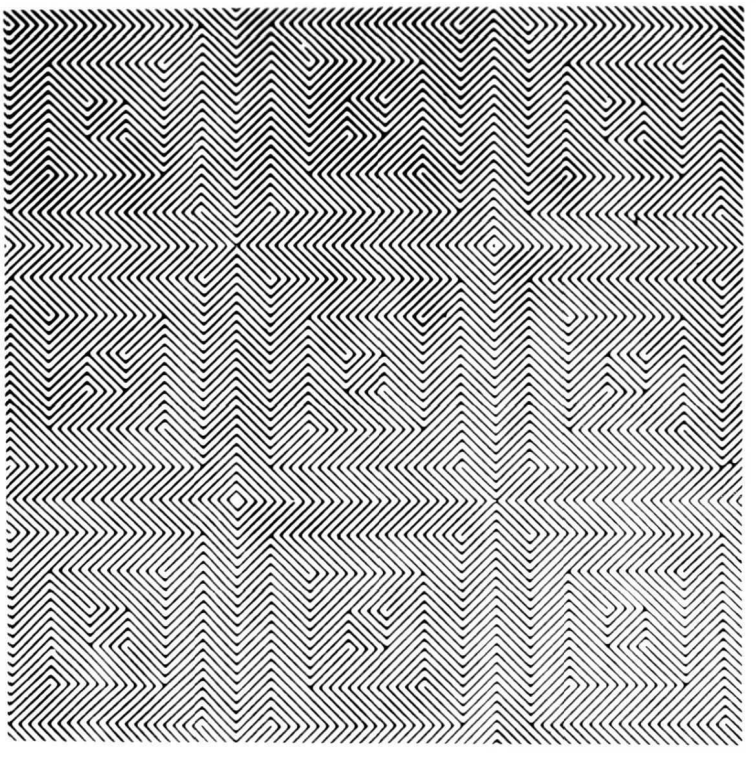


Figure 10. Detecting the motion of a moving random texture. See text for details.

(a)



(b)

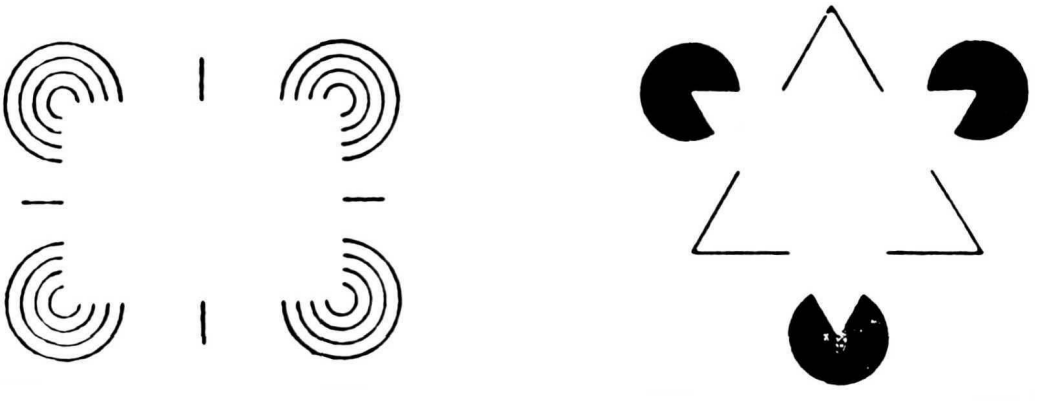


Figure 11. Contours that cannot be detected by simple intensity based operators.

SYMPOSIUM
AFFECT

Affect, emotion, and other cognitive curiosities

George Mandler
University of California, Dan Diego

The lure of phenomenocentrism. During the past century - at least since Darwin, Marx and Freud - our concept of reality has undergone changes that have been particularly apparent in the cognitive sciences. A dominant symptom has been the sharp swing of attitudes toward the apparent and convincing reality of phenomenal experience. It is practically impossible not to be overwhelmed by the immediacy of human experience. Philosophers have tended to accept its primacy through the ages and many cognitive scientists still do. We have overcome the perils of ethnocentrism and of anthropocentrism; we believe no longer that either our social values or humankind are the touchstones of social organization or the measure of animal behavior. But we are still phenomenocentric; we accept as basic the surface experiences that we seem to share with our fellow featherless bipeds. Phenomenocentrism has been abandoned in some corners of cognitive science; but many psychologists, philosophers and AI practitioners seem to hold stoutly to the dogma that common human experience provides the basic building blocks of cognition. I want to argue that this kind of commitment to surface structure as the psychologically "real" prevents progress and, in particular, has barred us from any reasonable understanding of human emotion. Any satisfactory comprehension of the structures and processes that generate affect and emotion requires the postulation of deep structure, a willingness to go "beyond phenomenology."

It is the case, of course, that the common experience our phenomenal world - does provide the major insights that have led us to determine the important problems and some of the answers about human cognition. Psychology has failed when it has ignored these insights. But the importance of common experience, language and understanding in the context of discovery must not color the need to go beyond phenomenology and folk experience when we try to understand the processes, structures and mechanisms that determine, generate and construct that experience.

There are two glaring examples of phenomenocentrism in the fields of affect and emotion. The first, popular through the ages and exemplified by Charles Darwin and his descendants, postulates the categories of emotion taken from the natural language as the fundamental building blocks of human (and lower) emotional life. It is a position that searches for the effects of unanalyzed emotions, posits a limited set of such fundamental emotions, and even seeks their locus in various corners of the human brain. It has variously reified such nicely vague concepts as fear, love, anxiety, joy, lust, dread, and so forth, and so forth.

The other position, beloved of poets and some social psychologists, posits the primacy and noncognitive (i.e., knowledge independent) nature of affective experience. Its advocates assert that feelings and affects occur prior to the registration and/or experience of other aspects of human experience, and particular prior to "cognitive" events.

Emotion and the categories of natural language. A concern that is related to phenomenocentrism and, like it, a hangover from the 19th century is exemplified in the penchant to take seriously the implicit claim that some natural language categories. define a well bounded, precise set of phenomena. Emotion is one of these categories. To ask "what is (an) emotion?" - as William James did and as some cognitive scientists still do - assumes that the vagueness and redundancy of natural language is suspendable. Categories like emotion (just as intelligence, justice, equity, learning, aggression, etc. etc.) have an evolutionary history and current function that do not support the weight of explanatory systems. They are useful, and have developed, as communicative devices in natural discourse. The analyses of these functions is an important enterprise (usually engaged in by our philosophical brethren). However, they fall far short of definitional devices as a first step toward satisfactory explanatory|y and theoretical ends. At least one can say that to date attempts to develop satisfactory, exhaustive, and scientifically useful definitions (much less explanations) of "EMOTION" have failed.

Again, having inveighed against natural language categories, I turn to them as a starting point. The categories of our common experience are, of course, collections of events (or objects) that do have a vague common core. That common core is - as in the case of the phenomenal experiences an indispensible starting point for serious investigation. In the case of affect and emotion, there are apparently two aspects that characterize the collection. Dictionaries tell us that emotions refer to "vehement, excited mental states," that they involve "agitation, disturbance of mind, feeling, passion." Affects are mental states that involve "desires, intentions," and "inward dispositions" and "intent." During its early history, and to some extent in its modern usage, the term "affect" also invoked the same kind of physical referent that the emotions do. What the common concepts of emotion and affect seem to have in common is a state of physi(ologi)cal excitation or arousal. What apparently differentiates the various affects or emotions are desires, intentions, and values.

Looking for emotion's deep structure. If we now turn to the problem of arriving at a program of theory or research, we need to postulate a system that constructs or generates some subset of these emotional phenomena. I shall defend one version of such a theory, but the main thrust of my argument is that some kind of theory (deep structure) needs to be developed that generates so-called emotional behavior and experience. Psychologists have variously emphasized the agitation/arousal dimension or the intention/value aspect. Most of the the papers at this symposium address the latter problem, i.e., the specification of the cognitive structure of emotions. I shall outline a point of view which is equally concerned with the arousal and evaluative aspects of the generative system, though I do not go into the kinds of details that the cognitive components require, and that are exemplified by the other papers of the symposium.

A point of view. My concern is specifically with the conscious experience of emotion. As a consequence, I have been concerned recently with the construction of conscious experience in general. There have been a number of recent suggestions, notably by A.J. Marcel (1981), that stress the constructive nature of consciousness, such that a particular conscious state is seen as constructed out of two or more activated schemas that produce a phenomenal unity that apparently conforms to the intentions of the individual and the requirements of the task and situation. The constructive approach to consciousness is ideally suited to accommodate the notion that conscious experience of emotion concatenates both evaluative

Copyright © 1981 George Mandler

cognitions and autonomic arousal. Thus, the phenomenal emotional experience is not some additive result of arousal and evaluation, but rather the schemas activated by arousal and evaluation are used in constructing the phenomenally unified emotional experience. Its intensity will indeed be related to the degree of arousal and its specific quality will depend on a complex evaluative cognitive event, but the two ingredients are experienced as a single emotion, just as eggs, milk, and sugar may be experienced as custard. This approach also accommodates the fact that there exist experiences of "cold" emotions, of evaluative cognitions without arousal, and of unemotional arousal, of autonomic arousal without the cognitions that provide emotional qualities.

The arousal component of emotional experience can be ascribed primarily to peripheral autonomic nervous system (ANS) activity. Whereas there is some evidence that what is most efficiently registered is some general level of ANS activity (heart rate, blood pressure, gastric motility etc), we also know that there are large individual differences in the patterning of the various autonomic indicators. It may well be the case that registration of peripheral arousal will, in the individual case, be governed by different patterns and may, in some cases, be driven by a single channel (such as heart rate activity). For the present, the sufficient and necessary conditions for the occurrence of autonomic arousal are not adequately known. To a large extent we still rely on lists of "elicitors" which are of varying degree of utility (e.g., tissue injury, stress, surprise, threat, emergency reaction, etc. etc.). I have suggested that the interruption of ongoing action, the discrepancy between expectation and evidence, and similar instances of incongruity between organisms' schemas and the evidence from the environment, are responsible for a large subset of the occurrences of ANS arousal. Such a hypothesis not only is consistent with the homeostatic view of the ANS, but also responds to adaptive, evolutionary functions of the autonomic system in general. Whenever an expectation is violated or a plan kept from being carried out (either in thought or action) an interruption (discrepancy) occurs which leads to ANS arousal. It is important to note that pleasant as well as unpleasant experiences are captured by this approach - unexpected, desirable events generate arousal just as unexpected, noxious events.

What is the nature of evaluative cognitions? In the first place, I suggest that there are three sources of value that influence the quality of an emotional experience. Our evaluations may be based on innate, prewired values - such as the preference for certain temperature ranges, the avoidance of looming objects, the preference for sweet over bitter tastes; or our values may be culturally predicated - we are "told" what is edible, lovable, drinkable, without ever having had direct interactions with the objects in question that would direct our values; and finally we may make valuative judgments that are determined by the structure of the valued event - or rather we base our judgment on some comparison between the event and some existing schema. It is the latter which I find most challenging; what are the structures that determine judgments of beauty, ugliness, preference or rejection, and which in turn determine emotions of joy and disgust, liking and disliking? There has been some reasonable amount of work done on such staples as anxiety, fear, and grief, but much analysis is still to come - some of it presented at this symposium. I have taken one step in that direction in trying to show how one of the more primitive evaluative judgment, that of preference arising out of the sense of familiarity, is related to the congruity between expectations (schemas) and evidence in the world.

In general, though, I would argue that much of the valuative cognitions that contribute to emotional quality deal with the internal structure of events rather than with the presence or absence of features or attributes. Thus, the sense of loss that leads to grief deals with relationships and not with the specific characteristics of the lost person. Similarly beauty and ugliness deal with structural relationships, as do the cognitions that underlie jealousy and even fear. That practically canonic emotion - anxiety - apparently has a cognitive basis in the perceived absence of structure, not in any definable feature of the anxiety arousing situations.

Emotion and cognitive science. In contrast to certain speculations to the contrary (e.g., Zajonc, 1980), these arguments suggest that evaluative cognitions (preferences, likings, aversions etc.) are relatively complex cognitive events, certainly involving more processing than simple definitional or featural judgments. We have recently collected some data that support this argument; simple impressionistic judgments of liking (the simplest evaluative judgment) are slower than simple categorical judgments, with the effect becoming rather large for familiar objects. Thus it takes longer to process the information needed for simple valuation than for simple categorization, exactly what we would expect if the former involves processing of internal structural relationships, while categorization may proceed on more simple presence/absence judgments about features or attributes.

The argument that affect or emotion is prior to or independent of cognitions frequently appeals to the phenomenal evidence that we are often conscious of valuations before we are aware of the details of the event that is being valued. Even if this particular kind of phenomenocentric assertion is confirmed by some future analyses, it does not say anything about cognition and affect, but it does address the nature of consciousness and the kinds of conscious constructions involved in affect and in other kinds of experiences. That analysis is beyond the scope of this presentation.

Cognitive scientists have, often for good reason, been accused of being scientific imperialists. The old division of the world into cognition, conation, and will has been destroyed by raiding parties that have penetrated deep into (undefended) territory previously considered noncognitive. To the extent that such aggression has been justified, it has been based only in part on the claim that all of experience and behavior deals with knowledge. More important may have been the fact that the new theory-rich cognitive science has been willing to take on all kinds of problems in terms of an information processing organism (animate or otherwise). For the time being, I have left aside the "cognitive" nature of autonomic arousal. However, in the spirit of cognitive imperialism one should be concerned with the kind of structural representations that would be useful theoretically and that would lead to a better understanding of the initiation, perception, and conscious construction of peripheral autonomic arousal. The neurosciences are

charter members of the cognitive sciences; maybe it is time to tell the peripheral physiologists to look to their borders.

References

Mandler, G. Mind and emotion. New York: Wiley, 1975.

Mandler, G. The structure of value: Accounting for taste. Technical Report No. 101, Center for Human Information Processing, University of California, San Diego, May 1981.

Marcel, A.J. Conscious and unconscious perception: II. An approach to consciousness. Cognitive Psychology, 1981, in press.

Zajonc, R.B. Feeling and thinking: Preferences need no inferences. American Psychologist, 1980, 35, 151-175.

Affect and Memory Representation

Wendy G. Lehnert

Department of Computer Science
Yale University

1. Affective Knowledge Structures and Computers

Many philosophers, psychologists, and A.I. workers have taken various positions on the issue of machines and emotion. Some argue that a computer can never "experience" human emotions in any significant sense because it just doesn't make sense to attribute conscicusness to an inorganic and programmed system [Gunderson 1971, Puccetti 1968, Scriven 1960, Ziff 1959] Others argue that our subjective sense of emotional experience is too "intuitive" and ill-defined a candidate for computational modelling [Dreyfus 1972, Weizenbaum 1976]. Still others argue that emotion will be a natural and necessary consequence of intelligent information processing, an inevitable side-effect of intelligence [Kenny 1963, Simon 1967, Doyle 1980, Sloman 1981]. And then there are always "rational purists" who consider emotional experience totally irrelevant to reasoning processes and therefore of no consequence to artificial intelligence whatsoever.

Many lines of reasoning have been invoked to secure these various positions [Dennett 1978], although most of the arguments are conducted with a distinctively philosophical tone. It is ironic that the most passionate advocates in these debates rarely argue from first-hand experience with computer simulations. Why does A.I. seem to be so silent on the the topic of emotion and computers? I cannot speak for everyone in the field, but I would guess that a lot of us prefer to avoid the whole morass because we believe that the questions being answered are not the questions we should be asking.

A computer can have knowledge of human emotionality in the same sense that it can have knowledge of mass spectroscopy, medical diagnostic techniques, or payroll data. Computers do not have to "be" emoting entities to use this knowledge any more than they have to "be" chemists, physicians, or bureaucrats to use knowledge specific to those professions. If it is difficult to give knowledge of emotions to computers, it is only difficult for the same reason that a thousand other topics are difficult for computers: people do not have a rigorous understanding of their intuitive knowledge in terms of information processing requirements. We need to develop (1) symbolic systems of internal representation, and (2) processing strategies to manipulate these symbolic structures. These two requirements are universal to all A.I. efforts, and the difficulties involved are not significantly amplified when the knowledge to be encoded is knowledge of human emotionality.

When we apply knowledge of emotions to an information processing task, we can evaluate our expertise on emotions by evaluating the overall effectiveness of the larger information processing system. What experience has A.I. had with affective knowledge structures? Our experience is admitedly limited, but it is not totally non-existent. To date, three distinct task orientations have touched on affective manipulations of one sort or another.

1) belief system maintainance
2) conversational simulations
3) narrative text processing

An ambitious implementation of belief system maintainance was attempted by Kenneth Colby in the early 60's [Colby 1967; 1973]. While Colby is best known for PARRY, the paranoid conversationalist [Colby 1975], his earlier work was aimed at a more general simulation of neurotic thought processes [Colby 1963; 1965]. Colby was specifically interested in simulations of Freudian defense mechanisms when they surface in clinical dialogues between psychiatrists and patients. His work involved affective manipulations, but only in a very superficial sense. Colby utilized "emotion monitors" which were numerical parameters with names like "excitation," "self-esteem," "danger," "well-being," and "pleasure." While Colby's simulations were never intended to implement a complete system of affective representation, he nevertheless found it necessary to maintain and manipulate these numeric parameters. For example, the "excitation" monitor reflected the overall anxiety of the system - a factor that any psychotherapist would want to take into account. Whether or not someone's anxiety level can be adequately represented on a scale of 1 to 10 is another question.

It is inevitable that belief system manipulations manifest themselves most naturally in interactive conversation. Colby was forced into conversational task orientations when he began his work on belief systems, and this eventually drew him toward PARRY. PARRY also utilized numeric parameters for "anger," "fear," and "mistrust," - a somewhat more narrow set than was needed for generally neurotic simulations. While PARRY is the only conversational system that I know of which has implemented an affective component, it is clear that any conversational system would require affective manipulations if it was designed to simulate emotional responses [Schank and Lehnert 1979].

Although Colby was primarily interested in thought processes, his simulations became thoroughly mired in language processing difficulties. Colby's sentence processing techniques relied on lexical pattern matching routines, and his internal memory representations were lexically-oriented as well. These devices were ineffective substitutes for natural language processing strategies, and Colby's models were significantly hampered by inadequate representational techniques [Boden 1977]. Similar impediments were encountered by other researchers who tackled belief systems early on [see e.g. Abelson 1973], so it comes as no surprise to see that the most recent work on belief systems is thoroughly grounded in theories of natural language processing and internal memory representation. [Carbonell 1978]. Whatever one's ultimate research goal (models of belief systems, memory organization, etc.) all dialog simulations are primarily natural language processing systems, and any attempt to circumvent this fact is destined to fail. In fact, when the research goal is a model of human emotion,

This work was supported in part by the Advanced Research Projects Agency under contract N00014-75-C-1111 and in part by the National Science Foundation under contract IST7918463.

it is far better to embrace the challanges of natural language processing with open arms: a natural language processing project provides the only naturalistic and realistic laboratory for theories of affective memory representation.

For example, narrative texts provide a rich proving ground for theories of affect - one need only look at stories that involve emotional reactions and emotional behavior. Yet affective knowledge representation has not been systematically tackled within natural language processing programs until very recently. The remainder of this paper will outline some of my own experiences in trying to implement affective knowledge structures in a narrative text processing system, BORIS.

The BORIS system currently utilizes a limited system of affective representation and affect-related inference mechanisms. In addition to these representational techniques for affects per se, the recently proposed TAU knowledge structure (Thematic Affect Unit) contributes to BORIS's affective inference capabilities as well [Dyer 1981]. Since descriptions of BORIS's processing techniques can be found elsewhere [Dyer and Lehnert 1980, Lehnert, Dyer, Johnson, Yang and Harley 1981], we will not go into a description of BORIS here - we will instead take this opportunity to explain how a computational model of affect might contribute to other models of affect that are not computational in nature.

To begin with, a computational model must address a specific set of problems to be solved. In language processing, affective manipulations are needed for four general inference situations:

(1) Given an event description, infer an affective antecedent:

"John took a valium." (=> John is upset)

(2) Given an event description, infer an affective consequent:

"John got a big raise." (=> John is happy)

(3) Given an affective state, infer its likely antecedent:

"After the hurricane, John was depressed."
(John suffered a loss => John is depressed)

(4) Given an affective state, infer its likely consequence:

John was so happy about his royalty check, he made reservations at Reno Sweeny's.
(... => John intends to celebrate)

In many cases, we must combine two or more of the above inference types to make sense of an implicit causality:

"After the hurricane, John saw a shrink."

To see how this process-orientation differs from a purely psychological approach to the problem, we will look at some inferences problems in a narrative text, and see how far a non-process-oriented model can go in helping a system like BORIS. When we first began to look at affect in BORIS, we were greatly inspired by Ira Roseman's model for representing affective states [Roseman 1979]. There are of course other approaches to affect [deRiveria 1977, Izard 1977], but we will not attempt to survey all the relevant proposals here. Readers who are familiar with alternative systems can judge for themselves whether similar troublespots would arise in trying to implement another system.

2. Conceptual Decomposition for Affective States

In the Roseman system, emotional states are represented by decomposition into five dimensions. Four of the dimensions assume positive and negative fields, while the fifth assumes a three-valued spectrum:

Five Dimensions of Affect

1) Motivational Status (desirability) [+,-]
2) Situational Status (attainment) [+,-]
3) Probability Status (certainty) [+,-]
4) Legitimacy Status (deservedness) [+,-]
5) Agency Status [self, other, circumstantial]

When an event is mapped into its appropriate place on each of the five spectrums, we can predict emotional reactions to the event. For example, (a) wanting a ticket to a sold-out Grateful Dead Concert describes a mental state with a positive motivation (wanting it) and a negative situation (not having it); (b) winning a ticket to a Grateful Dead Concert is an event with a positive motivation (wanting it) and a positive situation (having it); (c) losing the ticket has negative motivation (not wanting to lose it) and positive situation (having lost it); and (d) finding it again involves a negative motivation (not wanting it lost) with a negative situation (not having it lost). If all of this happens circumstantially, we expect to see (a) sorrow, (b) joy, (c) distress, and (d) relief. Using all five dimensions, Roseman's system differentiates 13 primary emotions. These are listed below with a vector encoding of the five dimensions as listed above. For example, (+ + + - S) corresponds to a positive motivation, positive situation, positive probability, negative legitimacy, and self-agency. An "*" indicates that the corresponding dimension can assume any value.

PRIMARY EMOTIONS

M S P L A		M S P L A	
(+ + + * C)	JOY	(+ + * * O)	LIKING
(+ + - * C)	HOPE	(- - * * O)	LIKING
(- - - * C)	HOPE	(+ - * - O)	DISLIKING
(- - + * C)	RELIEF	(- + * - O)	DISLIKING
(- + + - C)	DISTRESS	(- + * + O)	ANGER
(- + * + C)	FRUSTRATION	(+ * + O)	ANGER
(+ - * + C)	FRUSTRATION	(+ - - - C)	FEAR
(+ - + - C)	SORROW	(- + - C)	FEAR
(+ + * + S)	PRIDE	(+ - * + S)	REGRET
(- - * + S)	PRIDE	(- + * + S)	REGRET
(* * * - S)	GUILT		

Many lexical descriptions of emotionality are used to reference more than one conceptual configuration. For example, John could "regret" flunking a test (- + + + S). Alternatively, if John got a high B on the test, he might "regret" not getting an A (+ - + + S). These are two distinct senses of regret: we can regret what happened, and we can regret what didn't happen. People who dwell on (- + + + S) configurations kick themselves for past mistakes while people who dwell on situations involving (+ - + + S) are melancholy dreamers. We can describe either personality in terms of past

regrets, but important conceptual distinctions are lurking beneath these words.

Lexical ambiguities at the conceptual level make it difficult to describe Roseman's 13 primary emotions to everyone's satisfaction. For example, one could argue that "liking" is not an emotion at all but an attitude. The appropriate emotion for (+ + * * 0) and (- - * * 0) is really one of gratitude. Or perhaps "distress" should be called "discomfort." It is instructive to engage in this sort of criticism as an intuitive exercise, but a better way to test Roseman's system is with a computer implementation.

3. Implementing Affect

When we tried to implement Roseman's system in BORIS, we ran into some interesting difficulties. To get a sense of these, we will look at a sample text that BORIS processes, highlighting some problem areas encountered.

****

A BORIS Narrative

Richard hadn't heard from his old roommate Paul for years. Paul had loaned Richard money which was never paid back, but now he had no idea where to find his old friend. When a letter finally arrived from San Francisco, Richard was anxious to find out how Paul was.

Unfortunately, the news was not good. Paul's wife Sarah wanted a divorce. She also wanted the car, the house, the children, and alimony. Paul wanted the divorce, but he didn't want to see Sarah take everything he had. His salary from the state school system was very small. Not knowing who to turn to, he was hoping for a favor from the only lawyer he knew. Paul gave his home phone number in case Richard felt he could help.

Richard eagerly picked up the phone and dialed. After a brief conversation, Paul agreed to have lunch with him the next day. He sounded extremely relieved and grateful.

The next day, as Richard was driving to the restaurant, he barely avoided hitting an old man on the street. He felt extremely upset by the incident, and had three drinks at the restaurant. When Paul arrived, Richard was fairly drunk. After the food came, Richard spilled a cup of coffee on Paul. Paul seemed very annoyed by this so Richard offered to drive him home for a change of clothes.

When Paul walked into the bedroom and found Sarah with another man, he nearly had a heart attack. Then he realized what a blessing it was. With Richard there as a witness, Sarah's divorce case was shot. Richard congratulated Paul and suggested that they celebrate at dinner. Paul was eager to comply.

****

There are a number of important affect-related inferences in this story. For example, we should infer that Richard felt bad about spilling his coffee on Paul, and his offer to drive Paul home was motivated (at least in part) by a desire to alleviate guilt. In the next sentence, when Paul finds Sarah, we should not assume that Paul suffered a cardiac arrest; he is just very surprised. We must also understand why it made sense for Richard to congratulate Paul and suggest a celebration. What did Paul have to celebrate? Adulterous mates are not normally greeted with such enthusiasm, so the celebration must be causally connected to something else that Paul should feel good about. Notice that if Richard had expressed his heartfelt condolences to Paul instead of congratulating him, this would also make sense. Paul's affective state is complex and must be fully understood to accommodate these various possibilities.

To make affective inferences, BORIS needed to interpret events and states from the story in terms of Roseman's five affect dimensions. "Motivation" and "situation" were relatively easy to recognize by relying primarily on goal states. But the three remaining dimensions proved to be trickier than expected. We will look at some difficulties in "agency" recognition, although similar illustrations could have been chosen from "certainty" and "legitimacy" as well.

Initially, we thought that the agent for an event would simply be the physical actor of the event [Schank 1975]. We quickly discovered otherwise. For example, in a question answering task, experimental subjects indicated that Richard was happy to receive the letter from Paul. He wasn't grateful, and he didn't like Paul any more than before; he was simply happy. In order to infer that Richard was happy to get Paul's letter, we have to ascribe a circumstantial agent when Richard gets the letter. If the letter's arrival was encoded with other-agency, then Richard would either like Paul or feel grateful to Paul for getting the letter. Simple joy can only come from circumstantial agency. But the letter's arrival is encoded as an MTRANS event with actor = Paul (Paul sent the letter). If agency is not a function of an event's actor, what is it? BORIS was (and still is) stymied by the agency problem.

It seems that agency is a function of actors but more specifically, intentional actors. Notice how the affective inference changes if Richard believes that Paul sent the letter just because Paul wanted to make Richard happy. Now it is much more reasonable for Richard to like Paul or feel grateful to Paul for sending the letter. If X knowingly does Z to make Y happy, and Z succeeds in making Y happy, then Y will like X for doing it. If Paul does something intending to make Richard happy, then Richard experiences the event with other-agency. But if Paul does something which only makes Richard happy incidentally, then Richard experiences the event with circumstantial-agency. Knowledge of an actor's ultimate intentions is needed to establish affective agency for inference purposes.

In addition to intentionality, affective agency can be influenced by an actor's degree of social responsibility. For example, it makes sense that Paul got annoyed when Richard spilled coffee on him. But what is annoyance? Annoyance can be a variant of anger or dislike (Paul was annoyed with Richard), both of which require other-agency. Richard didn't intend to spill the coffee, but he was nevertheless responsible for the event (albeit innocently), and this responsibility gives us other-agency. Annoyance is even more ambiguous in the sense that it may also describe frustration, which involves circumstantial-agency: it is rotten luck to have someone spill coffee on you. If Paul is upset, but not upset with Richard specifically, then his annoyance is one of pure frustration.

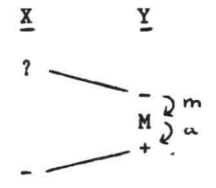
Since "annoyed" is ambiguous, and this particular example could go either way, it is useful to look for limiting cases which force one interpretation over another. For example, it seems reasonable that Paul might be more annoyed with Richard for the accident since Richard was drunk. If Richard were sober, he would somehow be less at fault. Suppose a frail little old lady is carrying a cup of coffee, and as she passes by Paul she collapses from a heart attack. Do we expect Paul to be angry at the old lady for spilling coffee on him? Not likely. Now suppose a boisterous drunk lurches past Paul and drops a drink on him. Do we expect Paul to be angry at the drunk? Sure. Neither event was intended, but a drunk is more responsible for his actions than a heart attack victim. People choose to get drunk, but they don't choose to have heart attacks. The element of free will operating in a drunk renders him more responsible for his accidents: a drunk chooses to be accident-prone. Since BORIS has no heuristics for assessing relative degrees of responsibility, BORIS defaults to circumstantial agency and therefore interprets Paul's annoyance as one of pure frustration. This is not altogether right, but a more correct interpretation requires an assessment mechanism for social responsibility.

One final problem with agency involves events that cause complex affect states. When Paul catches Sarah in their bedroom with another man, he witnesses and reacts to an event involving two actors. The event is assumed to be intentional (we are given no reason to interpret the bedroom activities as a rape) and at least one of the lovers must be responsible for it. So it seems that we have a clear-cut candidate for other-agency. Since this event will save Paul from a nasty court battle and divorce settlement, it is desirable from Paul's perspective. Sarah's activity can therefore interpreted by Paul as a desirable, positively attained, certain, and illegitimate (she's violating their marital contract) event of other-agency (+ + + - 0). But this configuration brings us to the improbable prediction that Paul will like Sarah and her lover, or feel grateful to them for engaging in their illicit activity.

The difficulty with this example is the complexity of Paul's emotional state. He may be happy about the settlement implications, but he is probably very unhappy about his territorial rights. Even if he doesn't feel possessive about Sarah (Paul did say that he also wanted a divorce), he has a right to feel put out by a stranger in his bedroom, to say nothing of his bed. His privacy is surely being violated on at least one level, and we are assured of his negative reaction when we are told that he "almost had a heart attack." So Paul's reaction is mixed: it has a strong negative component (- + + - 0) and a more far-sighted positive component (+ + + - 0) as well. This explains why it would make sense for Richard to either express sympathy or offer congratulations. It seems most appropriate to first offer condolences and then congratuations, but either one can be understood as a reasonable reaction on Richard's part.

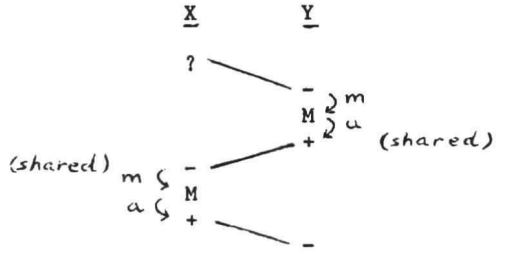
Special heuristics must be invoked for complex emotional states, and higher level knowledge structures will be needed to handle inferences in these cases. For example, the representational system of plot units (which grew out of our experience with BORIS's affect analysis), includes a special structure called "hidden blessing" to handle situations like Paul's reaction to Sarah [Lehnert 1980, 1981a, 1981b]. The hidden blessing plot unit encodes any event that causes an initial negative reaction which later yields to a dominately positive emotion. A similar "mixed blessing" plot unit handles cases where the initial reaction is positive, but a negative emotion follows.

In general, a plot unit is a configuration of three affect states: (1) "M" mental states with neutral affect, (2) "+" events that cause positive affects, and (3) "-" events that cause negative affects. Each affect state is interpreted with respect to a specific character, although plot units may contain multiple affect states that involve more than one character. For example, the retaliation unit involves two characters and five affect states:



This configuration tells us that X did something (?) which caused a negative reaction in Y (-). This negative event motivated a mental state (M) in Y, which was subsequently actualized by a positive event (+) for Y, and a negative event (-) for X. In other words, X did something that distressed Y, so Y retaliated by doing something to distress X. The vertical and diagonal links in this diagram describe various causal relationships between affects states within characters and across characters.

Affect state maps are constructed for each character in a story, as a way of tracking that character's emotional ups (+'s) and downs (-'s), and specific plot units are recognized when the linkages across affect states indicate that a given plot unit configuration has been encountered. If X were to get back at Y for Y's retaliation, we would have two instances of the retaliation plot unit sharing two common affect states:



A plot unit graph for a story can be generated by creating a graph node for each instantiated plot unit, and placing arcs between all pairs of plot units that share at least one common affect state. The above affect state map yields a graph of two nodes connected by an arc, while the BORIS divorce story involves 24 plot units in a connected graph structure.

It appears that the connectivity features in a plot unit graph provide a strong basis for summarization algorithms. When a story's plot unit graph contains a pivotal node (a node with maximal connectivity), we can expect a short summary of the story to be based on the conceptual content of that pivotal node. For example in the BORIS divorce story, the hidden blessing unit is pivotal, and we can summarize the story by saying "Paul saved himself from a nasty divorce settlement when he

accidentally found his wife in bed with another man." The hidden blessing unit encodes Paul's discovery as an event of mixed affect states, and a synopsis of any hidden blessing has to explain what was ultimately good about an initially negative event.

We have been experimenting with plot units primarily as a basis for narrative text summarization, [Lehnert 1980, 1981a, 1981b; Lehnert, Black, and Reiser 1981; Reiser, Lehnert, and Black 1981], but their use as a predictive knowledge structure for affective inference remains to be explored. Initial efforts in this direction led to the development of a slightly higher level of memory representation (Thematic Affect Units) that relates to adages and fables [Dyer 1981].

4. Conclusions

Our experience with Roseman's affect analysis and BORIS suggests that affective inferences are dependent on a substantial range of other inference mechanisms. It is not possible to study problems of affect without addressing seemingly unrelated problem areas. The current state of the art in language processing allows us to tackle recognition techniques involving scripts, goals, plans, interpersonal themes, plot units, and thematic affect units, all of which can contribute to affect recognition techniques. But affect analysis can also lead us into largely uncharted regions of intentionality and social responsibility, just to name two areas we've discussed.

We have not attempted to compile a list of all the related knowledge needed to handle affect, because this list is likely to be a comprehensive list of all knowledge structures used for language processing. Interestingly enough, there will probably be no knowledge structures devoted exclusively to affect. One could argue that the presence of a Roseman-like vector (+ - + + S) within computational memory constitutes an affect-specific knowledge structure. But this structure is just a processing artifact: we do not expect to find these vectors in long or short-term memory representations. If we needed to explicitly encode "John was angry," we might be reduced to vector notation, but as soon as this sentence is embedded in a context which tells us what happened to John and why he feels angry, his anger will likewise be embedded in some larger structure. TAU's fill this role already, and plot units (originally called "affect units") also operate along these lines. For example, vindictive or vengeful feelings can be readily derived from the retaliation plot unit.

In conclusion, I would say that it is altogether too early to pass judgement on any specific representational techniques for affects. Roseman has supplied the computational environment with a valuable framework in which to work (or perhaps play) and other psychological theories may also prove to be valuable, although I know of no others that have been (even partially) implemented. This particular area is a difficult one for A.I. implementations because it draws on so many other complex problem areas in memory and cognition.

It was nevertheless valuable to attempt an implementation of Roseman's model, if only for the sake of the spin-offs that emerged. By confronting problems of affect representation and affect-related inferences, we were led to the design of two new knowledge structures for narrative text analysis. Plot units and thematic affect units were natural devices for solving affect-oriented processing problems. The importance of affective knowledge structures for memory representation deserves further exploration in both the psychological and A.I. research paradigms. Recent research in both areas has uncovered some provocative results concerning emotion and memory which pave the way for further investigations. From the A.I. end of the world, we are seeing how the use of plot units for narrative text summarization suggests that emotional states of characters in a story play a far more central role in high-level memory representation than was previously suspected [Lehnert 1981a]. At the same time, psychologists are demonstrating how the moods and attitudes of experimental subjects can dramatically alter their patterns of memory retrieval [Bower 1981].

So affects appear to play many different roles in human cognition. In particular, we have seen how knowledge of emotional reactions can be crucial to various information processing tasks, including narrative text comprehension. Still a great deal of work remains to be done if we are going to fully understand the roles of emotional knowledge and experience in human information processing. We must turn to psychology labs and LISP programs for the answers to our questions, putting aside the more philosophical speculations about machines and emotions. It will be much easier to resolve our speculative debates in the face of some solid research, and the quality of our interdisciplinary dialogs will be greatly enhanced by empirical contributions from psychology and artificial intelligence.

BIBLIOGRAPHY

Abelson, R.P. (1973). "The Structure of Belief Systems." in Computer Models of Thought and Language. (eds: Schank and Colby). San Francisco: W.H. Freeman.

Boden, M. (1977). Artificial Intelligence and Natural Man. New York: Basic Books.

Bower, G. H. (1981). "Mood and Memory." American Psychologist. Vol. 36, No. 2. 129-148

Carbonell, J. G. (1978). "POLITICS: Automated Ideological Reasoning." Cognitive Science. Vol. 2, No. 1. 27-52.

Colby, K. M. (1963). "Computer Simulation of a Neurotic Process." in Computer Simulation of Personality: Frontier of Psychological Research (eds: Tomkins and Messick). New York: Wiley.

Colby, K. M. (1965). "Computer Simulation of Neurotic Processes." in Computers in Biomedical Research, Vol. 1 (eds: Stacy and Waxman). New York: Academic Press

Colby, K. M. (1967). "Computer Simulation and Change in Personal Belief Systems." Behavioral Science, 12, 248-253.

Colby, K. M. (1973). "Simulations of Belief Systems." in Computer Models of Thought and Language. (eds: Schank and Colby). San Francisco: W.H. Freeman.

Colby, K. M. (1975). Artificial Paranoia. New York: Pergamon.

Dennett, D. C. (1978). Brainstorms. Montgomery, Vermont: Bradford Books.

deRivera, J. (1977). A Structural Theory of the Emotions. New York: International Universities Press.

Doyle, J. (1980). "A Model for Deliberation, Action, and Introspection." AI-TR.581, MIT Artificial Intelligence Laboratory.

Dreyfus, H. L. (1972). What Computers Can't Do: A Critique of Artificial Reason. New York: Harper and Row.

Dyer, M. (1981). "The Role of TAU's in Narratives." Proceedings of the Third Annual Conference of the Cognitive Science Society. Berkeley, Calif.

Dyer, M. and Lehnert, W. (1980). "Organization and Search Processes for Narratives. Department of Computer Science Research Report #175. Yale University, New Haven, Conn.

Gunderson, K. (1971). Mentality and Machines. New York: Doubleday.

Izard, C. E. (1977). Human Emotions. New York: Plenum Press.

Kenny, A. (1963). Action, Emotion, and Will. London: Routledge and Kegan Paul.

Lehnert, W. (1980). "Narrative Text Summarization." Proceedings of the First Annual National Conference on Artificial Intelligence. Stanford, Calif.

Lehnert, W. (1981a). "Plot Units and Narrative Summarization." Cognitive Science. (in press).

Lehnert, W. (1981b). "Plot Units: A Narrative Summarization Strategy." in Strategies for Natural Language Processing. (eds: Lehnert and Ringle). New Jersey: Lawrence Erlbaum (in press).

Lehnert, W., Black, J., and Reiser, B. (1981). "Summarizing Narratives." Proceedings of the Seventh International Joint Conference on Artificial Intelligence. Vancouver, British Columbia.

Lehnert, W., Dyer, M., Johnson, P., Yang, C.J., and Harley, S. (1981). "BORIS -- An Experiment in In-Depth Understanding of Narratives." Department of Computer Science Research Report #188. Yale University, New Haven, Conn.

Puccetti, R. (1968). Persons: A Study of Possible Moral Agents in the Universe. London: Macmillan.

Reiser, B., Lehnert, W., and Black, J. (1981). "Recognizing Thematic Units in Narratives." Proceedings of the Third Annual Conference of the Cognitive Science Society. Berkeley, Calif.

Roseman, I. (1979). "Cognitive aspects of emotion and emotional behavior." presented at the meeting of the American Psychological Association, New York. (unpublished manuscript, Department of Psychology, Yale University, New Haven, Conn.)

Schank, R. (1975). Conceptual Information Processing. Amsterdam: North-Holland.

Schank, R. and Lehnert, W. (1979). "The Conceptual Content of Conversation." Proceedings of the Sixth International Joint Conference on Artificial Intelligence. Tokyo.

Scriven, M. (1960). "The compleat robot: a prolegomena to androidology." in Dimensions of Mind: A Symposium. (Ed: S. Hook). New York: New York Universities Press.

Simon, H. A. (1967). "Motivational and Emotional Controls of Cognition." (reprinted in Models of Thought. Yale University Press. 1979)

Sloman, A. (1981). "Why Robots Will Have Emotions." Proceedings of the Seventh International Joint Conference on Artificial Intelligence. Vancouver, British Columbia.

Weizenbaum, J. (1976). Computer Power and Human Reason. San Francisco: W.H. Freeman and Company.

Ziff, P. (1959). "The feelings of robots." Analysis. No. 19, 64-68.

SITUATION BASED EMOTION FRAMES AND THE CULTURAL CONSTRUCTION OF EMOTIONS

Catherine Lutz

1. Introduction: Anthropological Approaches

The question for this panel concerns the relevance of models of emotion for general theories of internal representation and processing. The particular question which I will address is that of the contributions which an anthropological or comparative approach can make to an understanding of emotional representations and emotional 'tasks'. In particular, I would like to suggest that we begin by looking at the way people themselves frame emotional experience and interactions. Much of this framing structure, moreover, is culturally provided and culturally variable. It can be seen in emotion word meanings, in the logic of emotional response, and in explicit ethnotheories of emotion.

Anthropologists have begun to document a diversity of theories of mind and self (e.g., Geertz, 1976; Leenhardt, 1979 (1947); Rosaldo, 1980; Strauss, 1977; White, 1981). Cultural knowledge systems, or 'ethnotheories,' explain why, when, and how emotion occurs, and they are embedded in more general cultural theories about the person, internal processes, and social action. The examination of these culturally constituted systems of knowledge can play two important roles in achieving the goals of cognitive science. In the first, we encounter ourselves in the other by seeing our own knowledge structures contrastingly highlighted. This is consonant with cognitive science's often stated aim of converting tacit understandings into explicit ones. Beliefs about the self are among the most tacit in any culture, as it is with them that intrapsychic and social reality are framed and felt. As Abelson (1979) has pointed out, one person's knowledge may appear to another as belief. Our 'knowledge' about emotion may turn out to be largely belief on cross-cultural inspection and comparison.

This is related to the second role for cross-cultural comparison in cognitive science which is that we may learn from, as well as about, the theories of other systems. For example, the emotion words of other cultural groups may have different referents or slice up the affective pie in different ways than do English emotion terms. If one of the aims of cognitive science is to develop an abstract, logical, and unambiguous language for a scientific psychology, the ethnosemantics of emotion will be an important topic of inquiry. A true science of affect needs a language which is transcultural.

Our most basic Western ideas about what constitutes humanness -- ideas which are dramatically evident in debates over whether computers can "feel emotion" -- form an implicit frame for our inquiries. The nature of this frame explains why emotion has been neglected, not only in cognitive science, but in social science generally. The traditional and still current concern has been with emotion as physical sensation. Internal feelings are presumed to be the primary referents of emotion words, and they are therefore seen as difficult to talk about -- emotions are, in this cultural framework, internal, private, pre-verbal states. Affect is not accidently omitted from the classic social science problem termed 'language and thought'.[1] Although it has more recently been acknowledged that one may think about how one feels and feel about how one thinks, the dichotomy of cognition and affect remains fundamental even in attempts to 'bridge' the gap between them.

We have value stances, moreover, towards the concepts and relations of our theories, and we are ambivalent about emotion. On the one hand, cognition is seen as 'higher' in evolutionary and other senses than emotion. Lakoff and Johnson (1980) present some striking evidence in English metaphors that emotion is so viewed. Metaphors which they term 'orientational' indicate that 'good', 'high status', 'control', and 'rational' are 'up', while 'bad', 'low status', 'being controlled', and 'emotional' are 'down' (1980: 14-17). The mature person ideally controls affect with thought. 'Intelligence', as that term is commonly defined in American English, refers to cognitive abilities, with 'emotional ability' being somewhat of a contradiction in terms.[2]

Emotion is instinctual, inchoate, formless, but at the same time it may also be positively valued. In scientific psychological terms, emotion is motivational and causes behavior. It is the life force, a sacred center. To claim emotion for computers, then, is sacrilege. For the above reasons, attempts at the 'cold' examination of a 'hot' topic (Abelson, 1963) appear to us, as Americans, both quixotic and profane.

I would propose that we proceed on the assumption that the terms 'cognition' and 'affect' represent, not dual and separate processes (Zajonc, 1980), but ideal types placed at the ends of a true continuum of more or less immediate and more or less motivated processing of events or situations. While emotion is the human potential for extremely high speed and action-oriented processing, emotions are culturally constructed concepts which point to clusters of situations typically calling for some kind of action. Many of these situation clusters will be universal while some will be environmentally specific. In the simplest example, one language group may code sudden events and dangerous events under the same rubric while another community might distinguish them. Those situations which will be universally framed together via emotion words will be those for which responses -- such as movement towards others or flight -- are shared and necessitated by the requisites of human social life. As D'Andrade has pointed out, a person who simply 'thought' coolly about acquiring food would have little chance of survival if that thought were easily extinguished and did not motivate remembering and action (1980: 16).

Such an affectless person could note that someone was in the process of stealing her car but, in the absence of anger, could be easily called off to play bingo.

Why do human communities need emotion words? It is not simply to name internal states. Emotion words are important primarily as communicators of one's perception of the occurrence of a particularly salient situation frame. The use of emotion words can often involve "backwards inferencing" on the part of listeners about the events which led up to the speaker's statement (White, 1979). In using emotion words, people also communicate the behavior which is likely to follow on their part. As one of the most crucial functions of social groups lies in their organization and regulation of individual behavior, emotion words are necessary for the most efficient and judicious coordination of plans, understanding, and behavior. By clustering situations which share action plans and other dimensions of meaning, emotion words facilitate such coordination both on the intra- and the inter-psychic levels.

2. Emotion Frames

The framing of emotion occurs at many levels. What has been and is here meant by a 'frame'? Interest in the framing of knowledge and social interaction has been widespread in social science. Attempts have been made to find person-centered, rather than investigator-originated, breaks in the stream of mental and social reality with a view to understanding how the psychologically and culturally constructed framed units organize human experience. These units have been a variety of cognitive ones (Minsky, 1975; Rumelhart and Ortony, 1976); they have been linguistic units which frame one or more concept, logical relation, or cultural proposition (Black and Metzger, 1965; Rosch, 1977); they have been situational units whose definition is a function of the transformation of "physical space and chronological time... into social space and social time" (McHugh, 1968: 3; see also Hall, 1977: 129-140); they may be contexts of interaction, the bracketing of which transforms an event into one with fundamentally different meanings and epistemological and behavioral entailments (Bateson, 1972: 177-193; Goffman, 1974); and on the most global level cultural groups are themselves framed by shared knowledge systems, customs, and world view -- an event which is framed in an East African culture may have very different meaning than that 'same' event framed in French culture. Cross-cultural and interpersonal understanding -- emotional and otherwise -- occurs only to the extent that there is a shared and agreed upon or a constructed frame for interaction.

There appears to be a growing recognition in psychology that the more inclusive frames are just that, rather than simply variables, and that the more micro-level frames are embedded in them. This is seen in the increasing concern with what is termed 'world knowledge'. It is also seen in the recent interest in metacognition (A. Brown, 1978; Flavell, 1976, 1978). Metacognition, or thinking about thought processes, is the ability to introspect about, monitor, and control those processes (Brown and Barclay 1976: 72). More basically, the term metacognition has been used to refer also to the awareness of the existence, nature, variables, and integration of cognitive processes (Wellman, n.d.). The importance of examining the phenomenon of metacognition lies in its potential role in the molding of more micro-level processes.

There is an important bridge to be built between these notions about metacognition and emerging anthropological concerns with ethnotheories of self and psychological process. Although work in metacognition has tended to be based on the assumption that these abilities and knowledge are culture-free, the child develops in a cultural milieu replete with explicit and implicit ideas about how one internally and socially operates. The child receives training in metacognitive skills in the context of that culture's typical learning settings. The psychologist's metacognitive skills and the anthropologist's ethnotheories thus constitute and constrain each other.

The concept of 'metacognition' should be seen, however, in the same ideal typical light in which we are here placing 'cognition', with 'metaemotion' coined to describe the more immediate and action-oriented processing of our processing. These metaskills can be conceptualized, like cognition and affect, as points along a continuum of delay and immediacy and of passivity and behavioral implication. To Brown and Barclay's list of metamemorial skills -- introspection, monitoring, and control (1976: 72) -- we must add, however, the evaluation of one's own cognitive and emotional processing. The skills of control and evaluation fall towards the same end of this continuum as does affect. People have knowledge concerning emotions and their place in the workings of the self and relationships, and they evaluate particular emotions (for example, 'righteous indignation is good', 'hate is bad', 'regret is good') and emotion in general. The inseparability of value and knowledge is seen in the fact that people in cultures which do not value a particular type of processing skill (such as intuition) will, according to the evidence accumulating in the fields of both metacognition and cross-cultural psychology, do less of such processing. Cultural values and knowledge determine what we term emotional experience, that is, they structure the way in which we frame events as salient, frame expressive communications from others as meaningful, and frame our own emotions linguistically and behaviorally.[3]

3. Ifaluk Theories of Emotion and Situation Frames.

The Ifaluk are a Micronesian people whose adaptation to their coral islet environment has been notably successful and resistent to colonial influences. Their one-half square mile atoll is densely populated, supporting 430 people on a diet extracted through fishing, horticulture, and gathering. Their society is organized

through the ranking of individuals based on their clan and lineage affiliations, gender, and age. Several chiefs head the matrilineal clans, and organize island-wide activities. The absence of physical aggression on Ifaluk marks it as one of the world's least violent societies. Ghosts of ancestral and other varieties are ubiquitous and dangerous, although they may also bring protection from typhoons and disease.

The Ifaluk have a rich body of knowledge about the nature of the person and emotion (Lutz, 1980, 1981). Although they do not distinguish emotion from thought, and do not have a monolexeme for either, they do label two closely related but nonetheless distinct types of internal processes, one more willful (tip-) than the other (nunuwan), more a product of individual desire than of social conformity or social intelligence. The Ifaluk words which we would call emotion words can be used to modify either of these two process terms.

The Ifaluk evaluation of emotion, that is, their metaemotional attitude, is a very positive one. Emotion is viewed as inextricably part of thought, and the correct understanding and behavioral enactments of emotions, both pleasant and unpleasant, are seen as signs of maturity. The mentally ill on Ifaluk are marked, in the indigenous view, by deficiencies in the ability to display both emotional and more technological understandings. Their view here interestingly converges with those of researchers recently working on the emotional-cognitive development of Down's syndrome infants (Cicchetti and Sroufe, 1976; Emde, Katz and Thorpe, 1978).

Emotion is a 'covert' category (C. Brown, 1974) for the Ifaluk. A domain of words exists which informants separate from others in sorting tasks. This domain includes words whose primary referents are organism-environment interactions. Ifaluk emotion words are defined by the situations in which people experience them. English emotion words, by contrast, foreground the organism response, with physiological signs (Davitz, 1969) taken as 'symptoms' of an internal event, rather than an external one. The two ethnotheoretical systems overlap to an important degree, however. While the Ifaluk have a less elaborate model of physiological correlates of events, some English emotion definitions point to the existence of a secondary situational model.

Joel Davitz, in his book The Language of Emotion, presents data on the meanings of emotion words for 50 American informants in a dictionary-like format. Data collection began in an open-ended manner, with people being asked to describe their experiences of each of 50 different emotions. A checklist was developed out of a large and representative sample of descriptive phrases. This checklist was presented to another 50 individuals to get comparable and quantifiable descriptors for the 50 emotions. The overwhelming emphasis on physiological state is evidenced, for example, in the definition of 'anger'. The following descriptions are listed in order of frequency (numbers in parentheses are percentages of the sample who checked each item).

> my blood pressure goes up...(72); I'm easily irritated...(64); I seem to be caught up and overwhelmed by the feeling (64); my face and mouth are tight, tense, hard (60); there is an excitement, a sense of being keyed up, overstimulated, supercharged (58); my pulse quickens (56); my body seems to speed up (54); there is a quickening of heartbeat (52); my fists are clenched (52) (Davitz, 1969: 35).

Contrast the meaning of 'anger' with the following definitions of the Ifaluk emotion word, song, which may be translated as 'justified anger'. "We are song from some gossip we hear [about ourselves which is not true], if someone talks a lot at us, [or] if the pig comes and eats food I just made." "If someone doesn't give me something I [legitimately] ask for, I'm song."

The evidence for the centrality of situation as a frame for Ifaluk emotional experience and understanding is found in several kinds of data. As just mentioned, folk definitions of emotion words are predominantly given in a form similar to the following, 'Emotion X is when someone steals my bananas', or 'If your child dies, then emotion Y'. Secondly, daily conversations abound in which emotion is explicitly discussed. Emotions of the self and others are spoken of in terms of their environmental or situational causes and correlates. Physical symptoms are not a salient topic of conversation. A third source of data is sorting tasks in which people were asked to pile-sort cards with Ifaluk emotion words on them. Post-test questioning of people about the rationale for their sorting decisions produces answers which most commonly cite a common eliciting situation or situation type, or alternately a sequence of situations (Lutz, 1982). Examples include, "These [emotion] words all go together because they all occur when we are interrupted in our work", and "They all involve something happening that we want [to happen]."

The Ifaluk logic of emotion is based on propositions about the parameters and meaning of commonly occurring situations. Why, then, and on what criteria do the Ifaluk frame something as a situation and frame diverse situations within the confines of a single emotion word? Let us return to the example of the word song, which has been translated as 'justified anger'. The prototypical situation which causes song is one in which another person has violated a cultural rule or value. Included are both situations where ego or a relative of ego suffers as a result and situations where it is simply cultural principle which is at issue, with no one directly disadvantaged as a result. Specifically, some of the commonly associated situations include hearing false gossip about oneself, encountering someone whose personality does not conform with the cultural values of generosity, calmness, even-temper, and respectfulness, being

excluded from the emotional and material support expected from kinspersons and others, and hearing of the violation of a taboo such as that against walking upright in front of the men's house.

'Anger' can also be defined by its situational correlates. Americans will define a frustrating situation as one in which anger typically occurs. 'Anger', according to Carroll Izard, arises in situations where one is

> either physically or psychologically restrained from doing what one intensely desires to do. The restraint may be in terms of physical barriers, rules and regulations, or one's own incapability (Izard, 1977: 329-330).

Brown and Herrnstein define 'anger' as the "illegitimate disappointment of legitimate expectations" (1975: 274). Although their definition is more restrictive than that of Izard, they are not contradictory. As Brown and Herrnstein point out, legitimacy is "neither universally acknowledged nor unchanging" (1975: 279). America is a poly-cultural and complex society, and there is much disagreement over values, rules, and regulations. This, and the value placed on individual achievement, make it unsurprising, therefore, that 'anger' may be evoked in situations where one's own individual goals (based on one's personal sense of the legitimacy of those goals) are blocked. The restraints and rules which Izard speaks of as eliciting 'anger' in Americans are seen as the correct and moral order on Ifaluk. A separate term, ngush, describes the state of being required to conform to a valid cultural rule which is nonetheless frustrating of other individual goals such as comfort.

Although the goals of individuals will be occasionally blocked in every culture, the definition of the situation in which one is blocked will vary from culture to culture, as will the subsequent emotional reaction (Whiting, 1944). The socialization process results in the molding of goals. The American child comes to expect certain kinds of achievement from her or himself. "Standing on your own two feet", "Making it", and "Proving yourself" are cultural aphorisms which reflect the types of independence and achievement goals which are acquired. The Ifaluk child, on the other hand, comes to operate with goals dictated by the values of sharing and interpersonal dependence. The behavioral implications of these values are outlined in cultural rules which provide for the equal distribution of everything from food to children and, conversely, for the equal distribution of sacrifice and restraint. The 'crazy' person on Ifaluk is partially defined as one who spends too much time alone, or thinks only of her or his own needs. Important differences in the nature, and hence the translations, of song and 'anger' flow from these differences in culturally constituted goals. It is for such reasons that American 'anger' results from "incapability" (Izard, 1977) while song results from the failure of the other to conform to group goals.

In attempting to compare emotion words across cultures, it can be seen that neither can the referent be assumed without examination of ethnopsychological theories in which they are embedded nor can the translation process be anything but primary. The translation of song as 'anger', or even as the more accurate 'justified anger', tends to erroneously suggest that the two terms share common referents and webs of ethnotheoretical meaning. A first step towards improving the accuracy of translation of emotion words might be to use situations rather than feeling tone as the primary criteria for mapping an emotion word in one language onto one or more emotion words in another language (Lutz, 1980).

Three final suggestions are in order about the general relationship between situation and emotion. Although the storage of emotional meaning in situation frames is particularly explicit in Ifaluk ethnotheory, the work of Minsky (1975) and others suggests that the situation might be a universal frame for many kinds of experience, as it represents an especially efficient way of storing related bits of information. The efficiency of this storage method might be enhanced, moreover, by certain types of metaemotional evaluations and controls and by developed metacognitive approaches to the situational coding of emotion.

Secondly, the function of emotion words may importantly consist of their linking of situations in a culturally meaningful way. 'Anger' links a certain group of situations, while song links another. Cultural adaptation calls for varieties of responses to the same 'objective' circumstances. Situations are correlated by their shared relationship to a particular cultural value and by the types of action which follow from them. Emotion words code these environmental correlations, and thus provide for understanding of self and social life.

Finally, the situations which are correlates and constituters of emotion in Ifaluk ethnopsychological theory are commonly occurring ones, such as confronting a task for which one is ill-prepared, observing a rule violation, or having one's child fall ill. These situation based emotion frames organize behavior through their links to action plans. Schank and Abelson claim that frequently encountered events usually require the development of accompanying scripts. These scripts are "highly stylized ways of executing planboxes" (1977: 96), while plans are the ranges of choices dictated by a particular goal. Thus we expect that Ifaluk cultural values, from which many individual goals originate, would contribute both to the framing of situations under a common emotion term, and to the available interactional script that dictates behavior in the emotion defined situation. We and the Ifaluk need emotion words to communicate definitions of the situation and intended plans of action. This and other ethnopsychological theories have much to teach us about the cultural

blinders which we take with us to the study of affect. Through the comparative investigation of emotion frames, including the linguistic, situational, and ethnotheoretic, we may be able to identify those frames which emerge only in particular environmental circumstances and those which are universally meaningful.

Notes

Acknowledgements. I would like to thank Geoffrey White for helpful comments on an earlier draft of this paper.

1. There are various forms in which the Whorfian hypothesis on the linguistic determination of thought may be construed from the strongest, in which non-verbal cognitive activity is patterned by language, to the weakest, in which only the perception of absent stimuli (i.e., memory) is strongly influenced (Miller and McNeil cited in Lemon, 1981: 202-203). If we include affect as part of the same continuum of internal processing on which thought is found (see below), the modeling of relationships between language and other processing will need to take account of both affect and cognition. We should expect that, under certain types of conditions of emotion activation, one or another version of the Whorfian hypothesis may apply.

2. For views of emotion as intelligence, see D'Andrade, 1980:15-17; Lutz and LeVine, n.d.; Meichenbaum, 1980: 274-278.

3. The sociologist Arlie Hochschild's (1979) notion of 'emotion work' is relevant here as a conceptualization of the way in which emotion knowledge structures ('feeling rules' in her scheme) affect emotional experience itself. In her view, these structures are provided by ideology and, in particular, by social 'feeling rules'. These rules govern not only behavior, but feelings themselves, and hence are not merely the 'display rules' of Ekman (1974).

References

ABELSON, R. Computer simulation of 'hot' cognition. In S. Tompkins and S. Messick (Eds.), Computer Simulation of Personality. New York: John Wiley, 1963.

ABELSON, R. Belief and knowledge systems. Cognitive Science, 1979, 3, 355-366.

BATESON, G. Steps to an Ecology of Mind. New York: Ballantine Books, 1972, pp. 177-193.

BLACK, M., and METZGER, D. Ethnographic description and the study of law. American Anthropologist, 1965, 67, 6(2), 141-165.

BROWN, A. Knowing when, where and how to remember: A problem of meta-cognition. In R. Glaser (Ed.), Advances in Instructional Psychology. Hillsdale, N. J.: Lawrence Erlbaum, 1978.

BROWN, A., and BARCLAY, C. The effects of training specific mnemonics on the meta-mnemonic efficiency of retarded children. Child Development, 1976, 47, 70-80.

BROWN, C. Unique beginners and covert categories in folk biological taxonomies. American Anthropologist, 1974, 76, 325-327.

BROWN, R., and HERRNSTEIN, R. Psychology. Boston: Little, Brown, 1975.

CICCHETTI, D., and SROUFE, L. A. The relationship between affective and cognitive development in Down's syndrome infants. Child Development, 1976, 47, 920-929.

D'ANDRADE, R. The cultural part of cognition. Address given to the Second Annual Cognitive Science Conference, New Haven, 1980.

DAVITZ, J. The Language of Emotion. New York: Academic Press, 1969.

EKMAN, P. Universal facial expressions of emotion. In R. LeVine (Ed.), Culture and Personality: Contemporary Readings. Chicago: Aldine, 1974.

EMDE, R., KATZ, E., and THORPE, J. Emotional expression in infancy: II. Early deviations in Down's syndrome. In M. Lewis and L. Rosenblum (Eds.), The Development of Affect. New York: Plenum Press, 1978.

FLAVELL, J. Metacognitive aspects of problem solving. In L. Resnick (Ed.), The Origins of Intelligence. Hillsdale, N. J.: Lawrence Erlbaum, 1976.

FLAVELL, J. Metacognitive development. In J. M. Scandura and C. J. Brainard (Eds.), Structural/Process Theories of Complex Human Behavior. Alphen a. 1. Ryn, The Netherlands: Sijthoff and Noordhoff, 1978.

GEERTZ, C. 'From the native's point of view': On the nature of anthropological understanding. In K. Basso and H. Selby (Eds.), Meaning in Anthropology. Albuquerque: University of New Mexico Press, 1976.

GOFFMAN, E. Frame Analysis: An Essay on the Organization of Experience. Cambridge: Harvard University Press, 1974.

HALL, E. T. Beyond Culture. New York: Anchor Press, 1977 (1976).

HOCHSCHILD, A. Emotion work, feeling rules, and social structure. American Journal of Sociology, 1979, 85, 551-595.

IZARD, C. E. Human Emotions. New York: Plenum Press, 1977.

LAKOFF, G., and JOHNSON, M. Metaphors We Live By. Chicago: University of Chicago Press, 1980.

LEENHARDT, M. Do Kamo. Chicago: University of Chicago Press, 1979 (1947).

LEMON, N. Language and learning: Some observations on the linguistic determination of cognitive processes. In B. Lloyd and J. Gay (Eds.), Universals of Human Thought: Some African Evidence. New York: Cambridge University Press, 1981.

LUTZ, C. Emotion words and emotional development on Ifaluk atoll. Ph.D. Dissertation, Harvard University, Cambridge, Mass., 1980.

LUTZ, C. Talking about 'our insides': Ifaluk conceptions of the self. Paper presented at the Tenth Annual Meeting of the Association for Social Anthropology in Oceania, San Diego, 1981.

LUTZ, C. The domain of emotion words on Ifaluk. American Ethnologist, 1982, 9, in press.

LUTZ, C., and LEVINE, R. Culture and intelligence in infancy: An ethnopsychological view. In M. Lewis (Ed.), The Origins of Intelligence. 2nd edition. New York: Plenum Press, forthcoming.

MCHUGH, P. Defining the Situation: The Organization of Meaning in Social Interaction. New York: Bobbs-Merrill, 1968.

MEICHENBAUM, D. A cognitive-behavioral perspective on intelligence. Intelligence, 1980, 4, 271-283.

MINSKY, M. L. A framework for representing knowledge. In P. H. Winston (Ed.), The Psychology of Computer Vision. New York: McGraw-Hill, 1975.

ROSALDO, M. Knowledge and Passion: Ilongot Notions of Self and Social Life. New York: Cambridge University Press, 1980.

ROSCH, E. Human categorization. In N. Warren (Ed.), Advances in Cross-Cultural Psychology. Vol. I. London: Academic Press, 1977.

RUMELHART, D. E., and ORTONY, A. The representation of knowledge in memory. In R. C. Anderson, R. Spiro, and W. E. Montague (Eds.), Schooling and the Acquisition of Knowledge. Hillsdale, N. J.: Lawrence Erlbaum, 1976.

SCHANK, R., and ABELSON, R. Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Systems. Hillsdale, N. J.: Lawrence Erlbaum, 1977.

STRAUSS, A. Northern Cheyenne ethnopsychology. Ethos, 1977, 5, 326-357.

WELLMAN, H. A child's theory of mind: The development of conceptions of cognition. In S. Yussen (Ed.), The Growth of Insight in the Child. New York: Academic Press, forthcoming.

WHITE, G. Some social uses of emotion language: A Melanesian example. Paper presented at the 78th Annual Meeting of the American Anthropological Association, Cincinnati, 1979.

WHITE, G. 'Person' and 'emotion' in A'ara ethnopsychology. Paper presented at the Tenth Annual Meeting of the Association for Social Anthropology in Oceania, San Diego, 1981.

WHITING, J. W. M. The frustration complex in Kwoma society. Man, 1944, 115, 140-144.

ZAJONC, R. B. Feeling and thinking: Preferences need no inferences. American Psychologist, 1980, 35, 151-175.

Disentangling the Affective Lexicon

Andrew Ortony and Gerald L. Clore

University of Illinois
at
Urbana-Champaign

When social psychologists and personality theorists investigate traits and emotions they frequently rely on lists of words that denote, or are thought to denote, traits, or emotions, or feelings. A classic problem in such work is that there is no unambiguous way of specifying which words refer to emotions, which refer to traits, and which refer to other behaviors, and non-emotional states. Investigators have generally relied on their intuitions in these matters, and by and large agreement has not been very high.

This problem is not so severe in lists of personality-trait words, largely because many modern empirical studies involving trait descriptors choose their terms from one of several "standardized" lists. For example, social psychologists frequently draw from the list of 555 words compiled by Anderson (1968). This list was developed by selecting feasible candidates from the 18,000 words appearing in Allport and Odbert's (1936) classic monograph. From the resulting reduced list of 2200 words, Anderson removed extreme words (e.g. majestic), words designating temporary states (e.g. aghast), words having to do with physical characteristics (e.g. hairy), strongly sex-linked words (e.g. alluring), and other words considered unsuitable as ingredients in impression formation (e.g. fond). Finally, words found to be unfamiliar to college students were eliminated. Although Anderson's list was also determined primarily on the basis of intuition, and although it does contain some ambiguous words (for example, happy, while certainly designating a trait, also can designate an emotion), it nevertheless has sufficient face validity to have gained wide acceptance.

Those who would study the emotions are less fortunate. Whether one seeks to map out the cognitive basis of the emotions, as we do, or whether one is investigating the effects of emotions on behavior, or of behavior on emotions, an indiscriminate use of language can be dangerously misleading in both theory construction and in the conduct of research. Many of the words in lists used in studying emotions either do not designate the kinds of states they are intended to, or they are ambiguous between different kinds of states. The indiscriminate use of such lists in theoretical and empirical research poses a serious methodological problem. For example, Russell (1980) scaled 28 "emotion-denoting adjectives". He found "sleepiness" to be an important dimension of such words. Although he lists the words used in his studies, he provides no justification for their inclusion and he describes no method for their selection. He included words like bored, tired, sleepy, drowsy, tranquil, and relaxed; we do not think that such terms denote emotions at all. If one includes among one's stimuli, words that have a high loading on sleepiness, sleepiness will turn out to be a factor. Until the inclusion of such words in the stimulus set can be justified, generalizations about the structure of the emotions have to be regarded as suspect.

The specification of necessary and sufficient criteria for emotions is a notoriously difficult if not an impossible goal. But because the employment of linguistic stimuli is an important avenue into the study of the emotions, some alternative method is needed for identifying emotion words. What we propose in this paper can be viewed as the linguistic groundwork required for language-based studies of the emotions and other affect-related terms.

The problem that we are dealing with is by no means unique to this domain. There are many areas where it is difficult or impossible to be entirely explicit about the criteria for class membership, but where psychologists have relied on rating scales and the intuitions of judges to achieve reliable and valid classifications. The work of Rosch and her colleagues (e.g. Rosch, 1978; Rosch & Mervis, 1975) on the categorization of concrete objects is an obvious example.

In the present context, one possibility would be to present subjects with candidate emotion words and ask how good good they are as examples of emotions. This we plan to do. However, as the sole strategy, this approach has drawbacks related to not knowing what criteria subjects employ in their judgments, and consequently it raises troublesome problems about reliability. Thus, as a first step, we chose to employ a number of explicit tests that we hope offer greater reliability and that provide potentially useful additional information about the structure of the affective lexicon. These tests constitute a set of heuristics for isolating genuine emotion words (and other kinds of words) from a list of putative emotion words. They take the form of a group of sentence frames into which a candidate word is inserted. The tests are "passed" or "failed" by a particular word depending on the extent to which groups of judges consider the resulting sentences to be meaningful rather than anomalous.

Finally, it should be emphasized that we think of these tests as a set of heuristics or "rules of thumb" rather than as an algorithm. Nevertheless, we think that they do a tolerable job of disentangling the affective lexicon--certainly a better job than blind intuition, or than no criteria at all.

While our primary goal is to isolate the genuine emotion words from a pool of candidate emotion words, we also consider it interesting to attempt to classify the major kinds of words appearing in the pool. The pool comprises the union of several lists claiming to be lists of emotions and/or of feelings. In constructing it we drew primarily from lists used in various psychological treatments of emotion (Bush, 1972; Dahl & Stengel, 1978; Davitz, 1969; Russell, 1980). The final pool consisted of about 500 words, the largest contribution coming from Dahl and Stengel's extensive list. A sizeable number also came from Bush (1972), who had reduced a prior list of 2,186 adjectives from Allport and Odbert (1936). From these Bush selected the 263 words that raters agreed were more relevant to emotions ("what a person feels") than to personality ("what a person is like") or to behavior ("what a person does"). Also included was Davitz's list of words from Roget's Thesaurus, and other smaller lists.

We found as we examined these lists that while we could not give a satisfactory definition of an emotion, we could readily eliminate many of the candidates as words that did not refer to emotions. For example, in the lists of words designating emotions and feelings used by Dahl and Stengel (1978) or Bush (1972), there are numerous "intruders" such as tired, hungry, breathless, and revived-- words which seem to designate body states, and words like confused, baffled, and sure, which seem to represent

non-affective cognitive states. Still other entries like abandoned, abused, and appreciated represent the acts or beliefs of others relevant to the self; they could certainly cause emotions but do not themselves denote emotions.

The linguistic tests that we propose are attempts to classify such "intruders" in a reasonably systematic way while also separating out emotion words. The first distinction we make is between words that designate traits or emotions and words that do not. Words that do designate traits or emotions are of three kinds: (a) "pure" trait words, which refer only to traits (and not to emotions), (e.g. studious, ambitious, mean), (b) "pure" emotion words, which refer only to emotions (and not to traits) (e.g. jubilant, distressed, embarrassed), and (c) polysemous words that can be used to refer to both emotions and traits (e.g. cheerful, happy, proud). For brevity we shall refer to such words as "emotion-trait hybrids". Although less central to our concerns, still of interest are the three kinds of words already mentioned that constitute the other half of the pool. These we call "body-state" words, "cognitive-state" words, and "other-action" words. The tests that we discuss in this paper are all designed to deal with adjectives or participial forms. Rephrasing of the tests is required to handle noun and verb forms.

Words denoting emotions or traits

The first test that we propose is actually a pair of sentence frames. One, Frame A, deals with negatively valenced words, and one, Frame B, deals with positively valenced ones. These frames can be thought of as linguistic filters. Their logic is to contrast candidate words with something explicitly emotional so that words that do not denote emotions will produce meaningful (as opposed to anomalous) sentences. The test separates the entire pool into two halves: (a) an item that fails the test (i.e. produces an anomalous sentence) is most probably a word that denotes a trait or an emotion, and, (b) an item that passes the test (i.e. produces an acceptable sentence) is probably a body-state word, an other-action word, or a cognitive-state word. Thus, the test is intended to allow as sensible completions only words like puzzled and certain (cognitive-state words), breathless and refreshed (body-state words), and abandoned and appreciated (other-action words).

Test 1.
Frame A: Although at that moment Mary was xxxxx, she was emotionally content
Frame B: Although at that moment Mary was xxxxx, she was not emotionally content

The word although anticipates a contrast, and in the contexts of these frames, it is a contrast of valence. However, the presence of the phrase emotionally content, constrains the contrast to non-emotional terms.

Accordingly, emotion words will fail the test, but body-state words, cognitive-state words, and other-action words all pass it. For example, words like breathless, puzzled, and abandoned pass the test because they fit the sentence frame for negative words (Frame A), and words like refreshed, certain, and appreciated pass because they fit the frame for positive words (Frame B). Traits are prevented from fitting into the sentence frames by incorporating in the frames a reference to a particular moment so that a quality that is enduring will give rise to an anomalous sentence. Thus, trait words as well as emotion words fail the test (e.g. honest, unkind, jubilant, and distressed).

Since our primary goal is to separate trait descriptors from emotion words, we shall deal first with that part of the initial pool that fails Test 1. Recall, first, that terms like proud, sad, and happy are sometimes used as trait descriptors and sometimes as emotion words. Thus, the half of the pool containing traits and emotions actually contains words of three kinds--the "pure" emotion words that unambiguously designate emotions (e.g. embarrassed, disgusted, jubilant), the "pure" trait words that unambiguously designate traits (e.g. thrifty, intelligent, studious, dishonest), and the "emotion-trait hybrid" words that have two senses, one referring to an emotion and one to a trait.

The test that we now describe is designed to separate pure trait terms and emotion-trait hybrids from pure emotion terms. Because the context provided by the sentence frame resists temporary states in favor of perservering qualities, it allows as sensible completions only traits and hybrid words with a trait as one meaning.

Test 2: John was well-known as a(n) xxxxx person

The result of applying this test is to separate examples like the following:

(PASS)	(FAIL)
pure traits and hybrids	pure emotions
anxious	disgusted
happy	distressed
proud	embarrassed
materialistic	jubilant
superstitious	love-sick

In order to separate the hybrids from the pure traits, another test, Test 3, is needed. This test may be applied to the same set of words as Test 2. The hybrids can then be isolated by taking the intersection of words passing Test 2 and of those passing Test 3. This is because Test 2 detects words that have trait readings, while Test 3 detects that subset of them that also have emotion readings (see Fig. 1).

The rationale behind Test 3 is that emotions can be experienced to varying degrees, and can be experienced in the absence of an interpersonal exchange. Thus, reflecting on a situation can give rise to an emotion but not to a trait, although, if a term is ambiguous as between a trait and an emotion, it will fit the test because of its emotion sense.

Test 3: As he reflected on what had happened, John was quite xxxxx

The result of applying this test is to separate examples like the following:

(PASS)	(FAIL)
pure emotions and hybrids	pure traits
cheerful	ambitious
distressed	intelligent
disgusted	knowledgeable
ecstatic	mean
frightened	sensitive
proud	thrifty

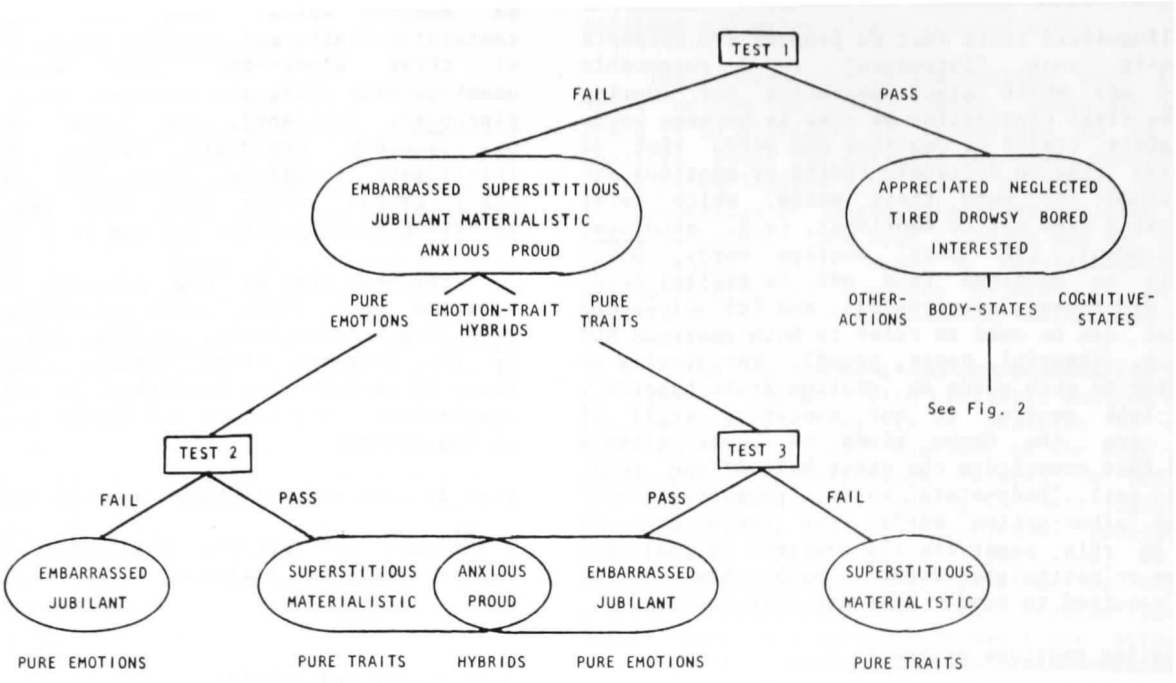


FIGURE 1.

Our primary goal has now been achieved. We think we have proposed a reasonably methodical procedure for isolating emotion words from a pool that contains some words that do not denote emotions. Furthermore, we have separated two kinds of emotion words, those that seem to denote emotions exclusively, and those (hybrids) that also denote traits. We consider this to be a potentially important distinction. The results of multidimensional scaling studies, for example, in which subjects make similarity judgments can be muddied by the unwitting inclusion of a subset of ambiguous stimuli (i.e. hybrids).

Words not denoting emotions or traits

There may be occasions on which one might want to compare emotion words to some other kinds of words, say, cognitive-state words, or other-action words. Although of secondary interest to us, the same kind of procedures can be used to separate the three kinds of words appearing in the other half of the initial pool, namely the half comprising words that passed Test 1 (see Fig. 2).

The first kind of words that we attempt to isolate are those that do not represent an internal state of a person at all. These we call "other-action" words because they characterize the actions (or the attitudes) of others that are relevant to (although not necessarily directed towards) the self. Perhaps because they are so strongly associated with emotional responses many other-action words have found their way into lists of emotions and traits. For example, the word abandoned appears in the lists of Dahl and Stengel (1978) and of Bush (1973). It also appears in the Personal Traits column of Allport and Odbert's (1936) list. However, in modern English abandoned designates neither a trait nor an emotion. One cannot be disposed to behave "abandonedly", (although we do speak of behaving "with gay abandon" meaning wrecklessly), and one does not experience "being abandoned" as a separate emotion. Rather, abandoned represents the actions of some other vis a vis the self. Its special quality, its emotional loading, presumably comes from the fact that the knowledge that one has been abandoned typically gives rise to (negatively toned) emotions. It is, however, as much of an error to assume that "abandoned" is an emotion or a trait as it would be to suppose that "kicked in the groin" was.

We shall assume that unlike emotion words, the most salient characteristic of other-action words is that others can engage in those actions without the person to whom they are relevant necessarily being aware of them. Since one can be abandoned and not know it, abandoned cannot represent an emotion or any other kind of internal state; it is an other-action word. Notice that it does not follow from this that awareness entails an emotional state. Normally, awareness is a necessary but not sufficient condition for an emotion. Thus, Test 4 is designed to identify other-action words. The logic of the test is (a) to deny awareness by using the expression "totally unaware", and (b) to take advantage of the fact that "other-action words" require actions by others that might influence the self by explicitly making the agent of the action an other.

Test 4: John was totally unaware that he
had been xxxxx by the woman

For this sentence frame reasonable completions are restricted to words that denote the actions (physical or mental) of others. As with the earlier tests, some words will fail to fit simply because they are of the wrong syntactic type, but, as always, the more interesting cases are those for which the resulting sentence is not syntactically ill-formed but

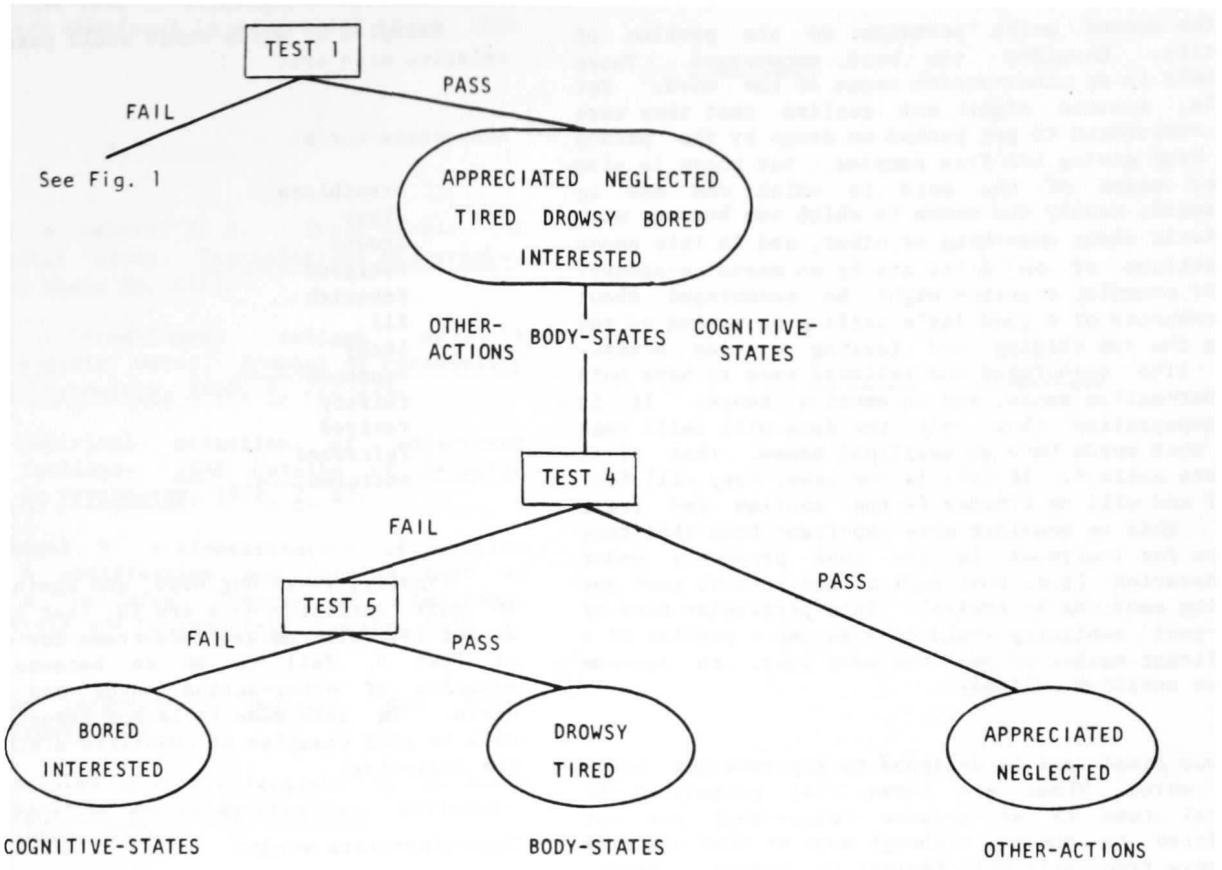


FIGURE 2.

rather is semantically anomalous. For example, the word puzzled does not fit very well because it is odd to suppose that John could be puzzled and not realize it. We think that puzzled fits better into the category that we call "cognitive-state" words. A more difficult example is revived. Because revived suggests the possibility of prior unconsciousness it seems better able to fit into the sentence frame, yet we like to think that revived is a body-state word. If this is so, then subjects should judge it to fit better in the body-state frame (see Test 5, below), even though it might do tolerably well in the other-action frame. Since subjects are asked to judge how well a target word fits in a frame, it would be sufficient for our purposes to discover that words like abandoned and ignored fit better than words like revived. It is not necessary that words not in the category upon which we are focussing give rise to seriously anomalous sentences. Our expectation is only that the most reasonable completions are produced by other-action words. Internal state words make poor completions. Thus, we can separate other-action words using this test. When subjects in our experiments are instructed to make meaningfulness judgments they are warned to ignore one particular interpretation of the sentence-frame in Test 4 that would confound the results by rendering spurious "meaningful" judgments. Subjects are told that the focus of the sentence should be on John's lack of awareness, not on the identity of the agent responsible for the action. Thus, they are instructed to ignore the interpretation in which John might have wrongly attributed his being ignored, appreciated, revived etc. to someone other than the woman. The following are examples of words that we think pass Test 4 most easily.

Other-action words:

abused
abandoned
appreciated
defeated
disgraced
ignored
neglected
slighted
welcome

It is worth pointing out a couple of things at this juncture. The first concerns the difference between "being" and "feeling" something. The inclination to treat other-action words as internal state words is much greater when they occur with "feel" than with "is". The reason is that "feel" can be, and often is interpreted to mean "feel as one would if (one realized that) one was xxxxx". Thus, that John was ignored entails nothing about what John felt. It merely asserts that somebody ignored John. Whether or not John responded to this other-action emotionally will depend on all kinds of factors (e.g. Was he aware of the fact? Did he expect anything else? Did he care? etc.) In other words, the inference to an emotional response is a pragmatic one, not a logical one. Yet, if one says that "John felt ignored", we have much more license to infer that John was in an (emotional) internal state. We infer that John responded emotionally. There is no doubt that feeling ignored is a unique kind of (negative) feeling--so too is the feeling of being pricked by a needle. But this fact is not sufficient for it to

count as an emotion. One cannot confuse causes with their highly correlated effects.

The second point pertains to the problem of ambiguity. Consider the word encouraged. There certainly is an other-action sense of the word. For example, someone might not realize that they were being encouraged to get hooked on drugs by the person that kept giving him free samples. But there is also another sense of the word in which one can be encouraged, namely the sense in which one becomes more optimistic about something or other, and in this sense the actions of an other are by no means necessary. So, for example, a person might be encouraged about the prospects of a good day's sailing on waking up and seeing the sun shining and feeling a fine breeze. Words like encouraged and relieved seem to have both an other-action sense, and an emotion sense. It is our expectation (but only the data will tell) that where such words have an emotional sense, that sense is more salient. If this is the case, they will fail Test 1 and will be treated in the emotion and trait pool. This we consider more important than that they survive for treatment in the pool presently under consideration (i.e. that part of the initial pool not denoting emotions or traits). This particular kind of cross-pool ambiguity would only become a problem if a significant number of emotions were lost, an outcome that we consider unlikely.

Our final test is designed to separate out body-state words. These are terms that pertain to the physical state of an animate being--they are not restricted to humans, although some of them might be used more frequently with respect to humans. Again, these words can be valenced, and are often, but by no means necessarily, associated with emotional responses. Their appearance in various lists of emotion words (e.g., again, those of Dahl & Stengel, and of Bush, and of Russell) is probably due to the fact that they appear in (Column II of) the Allport and Odbert (1936) list. This category is loosely characterized by these authors as "terms designating mood, emotional activity, or causal and temporary forms of conduct" (p.vii). In it appear words like thirsty and breathless which in our opinion do not fit even this loose characterization. What is it to be in a thirsty mood? Is being thirsty an emotional activity, or a form of conduct? We suspect that these terms appear in Column II not because they belong there, but because they are less incongruous there than in one of the other three categories used by Allport and Odbert.

Test 5 attempts to separate out these terms by using a sentence frame that focuses on body states (as opposed to other kinds of sensations, or perceptions), and that minimizes the cognitive content by predicating them of a newborn infant:

Test 5: The pediatrician explained that one of the physical characteristics of a newborn infant was to be xxxxx.

It seems to us that this test only allows as good completions terms that designate body feelings. It seems to us to more readily allow completions with words that do not entail awareness and that do not suggest cognitive activity (emotional or otherwise). Thus it would be odd to complete this frame with an other-action word like ignored, and it would be odd with words like certain. The oddness arises both from attempting to predicate higher level cognitive functions involving social awareness and metacognition to newborn infants, and from the fact that these predicates do not refer to physical characteristics.

Examples of words which would pass this test with relative ease are:

Body-state words:

breathless
dizzy
drowsy
fatigued
feverish
ill
itchy
nauseous
thirsty
revived
refreshed
satiated

Finally, it is our hope, and again we shall have to wait until the data are in, that those words that do not fit well the sentence frame for either Test 4 or Test 5, fail to do so because they are poor examples of other-action words and of body-state words. In that case it is our expectation that they will be good examples of cognitive-state words such as the following:

Cognitive-state words:

bored
disbelieving
distracted
doubtful
overworked
puzzled
uncertain
uninspired
amused
aware
certain
impressed
interested
sure
vindicated

Conclusion

The question of the psychological validity of the various categories that we have proposed is obviously an important issue. We find these categories to be intuitively reasonable and we believe that they do represent psychologically important distinctions. However, ultimately we would like to know that these distinctions correlate with behavioral differences. For example, in a pilot study conducted by Lord and Ortony memory for emotion words was found to be very much superior to memory for cognitive-state words. These are the kind of data that one needs to demonstrate the psychological validity of the distinctions we have proposed.

Finally, we should point out that we are more wedded to the general principles that we have proposed than we are to the specific tests. Indeed, some of the tests we find rather inelegant. It remains to be seen how effective these tests are, and we are convinced that there is room for considerable improvement. However, some procedure along the lines of the one we have suggested seems essential if one is to avoid the kinds of problems in the analysis of

emotions that we identified at the outset. We hope that our discussion will alert those who are studying the emotions to the need to distinguish between states that genuinely are emotional in nature and those that are not.

References

Allport, G. W., & Odbert, H. S. Trait-names: A psycho-lexical study. Psychological Monographs, 1936, 47(1, Whole No. 211).

Anderson, N. H. Likeableness ratings of 55 personality-trait words. Journal of Personality and Social Psychology, 1968, 9, 272-279.

Bush, L. E. Empirical selection of adjectives denoting feelings. JSAS Catalog of Selected Documents in Psychology, 1972, 2, 67.

Dahl, H., & Stengel, B. A classification of emotion words: A modification and partial test of de Rivera's Decision Theory of Emotions. Psychoanalysis and Contemporary Thought, 1978, 1, 269-312.

Davitz, J. The language of emotion. New York: Academic Press, 1969.

Rosch, E. Principles of categorization. In E. Rosch (Ed.), Cognition and categorization. Hillsdale, N.J.: Erlbaum, 1978.

Rosch, E., & Mervis, C. B. Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 1975, 7, 573-605.

Russell, J. A. A circumplex model of affect. Journal of Personality and Social Psychology, 1980, 39, 1161-1178.

Acknowledgements

We would like to thank Charles Lord and George Miller for helpful comments on earlier drafts of the paper. The research reported herein was supported in part by the National Institute of Education under Contract No. HEW-NIE-C-400-76-0116, and in part by a Spencer Fellowship awarded to the first author by the National Academy of Education.

Affect amd the Perception of Risk

Amos Tversky and Eric J. Johnson
Stanford University

As cognitive scientists turn their attention to emotion, they face the task of integrating affect into models of cognition. The perception of risks seems an ideal area to examine the relationship between cognitive and affective processes. When we witness an accident, or read a newspaper story about a natural disaster, we do more than simply revise our subjective probabilities. We are often quite disturbed and shaken by such events. Our encounters with risk are inevitably connected with feelings, including those of surprise, dismay, and worry.

Previous work in risk perception has concentrated on the cognitive domain. Lichtenstein et al. (1978), for example have asked people to estimate the frequency of death due to various causes. They argue that the availability of instances in memory helps determine these perceived frequencies. Thus, homicide is seen as much more common than suicide, although actually the reverse is true. Causes of death which are spectacular and the subject of media coverage appear to be overestimated while more mundane causes are underestimated.

We conducted three studies using an experimental paradigm similar to the one used by Lichtenstein et al. Before they made their estimates, however, subjects read a newspaper-like account of the death of a single individual under the guise of a newspaper reporting study. These stories, although quite graphic, were relatively devoid of information. They were, however, effective in changing mood, causing readers to report they felt much more depressed than a control group which had not read the stories. Later, in an apparently unrelated questionnaire, these subjects were asked to estimate the frequency of death due to various causes. The causes of death ranged from those closely related to the topic of the story, such as stomach and lung cancer for a story about leukemia, to unrelated causes such as tornados and airplane accidents.

The potential impact of these stories, and their accompanying changes in mood, represent a continuum. At one end of the continuum, we might expect the story to have no effect on the estimates. This is the normatively justified response, since the stories contained no information about the frequency of the death in the population. In contrast, the reader of the story might generalize from the instance in the newspaper-like story and increase their estimate of the frequency of that cause of death. We will term this a local generalization.

The impact of the story might also generalize to other, related risks. A story about a leukemia victim might also raise our subjective probability of related diseases such as lung and stomach cancer, but not unrelated risks such as airplane accidents. This gradient generalization should be closely related to the similarity of the risks. Finally there is abundant evidence in social psychology (Isen, Shalker, Clark, and Karp 1978) for more pervasive influences of affect. We might expect that increases in estimated frequency might occur for all risks, a possiblity we term global generalization.

In the first two studies we examined the generalization of negative affect across the responses. Despite our attempts to provide a sensitive test of local or gradient generalization, both studies demonstrate sizable global generalization. Readers of the newspaper stories estimated that all causes of death were about 40% more common than the control. Since the changes were unrelated to the topic of the story, these data suggested that the effect was due to mood induced, and that the bad moods were more than unpleasant states. In addition, they had pervasive influences on an important class of risk-related judgments.

In the third experiment, we broadened the estimates we requested to include items not related to either death or risk. For example, subjects were asked to report the frequency of bankruptcy and divorce. Even with these widely divergent estimates, we have found strong global generalization of affect, with no evidence for either local or gradient generalization. We also included a condition which read an additional newspaper story free of risk related content, but which described a series of negative events which occurred to the main character. Since the story made no reference to risk or death, its principle effect was the negative mood it induced in the reader. This depressing story resulted in a pattern of results almost identical to those induced by the risk-related newspaper stories.

These data, viewed as a whole, demonstrate that affect can have a large and pervasive influence on one important class of judgments, estimates of the frequency of risk-related events in the population. So far we have found no indication of a connection between the information contained in a story and its impact on the estimated frequency of death. The overriding factor in these increases does not appear to be the story told, but rather the mood or affect state it conveys to the reader. These effects are not limited to areas of death, but have been shown for estimates of non-fatal hazards and lifestyle threatening risks such as divorce and bankruptcy.

Any model of affect must account for two important aspects of this phenomenon: (1) Induction of a negative mood alone is sufficient to change estimates, and (2) the size of the change is unrelated to the semantic similarity, either among the estimates themselves, or between the cause of the mood and the estimates.

References

Isen, A. M., Shalker, T.E., Clark, M. & Karp, L. "Affect, accessibility of material in memory, and behavior: A cognitive loop?", Journal of Personality and Social Psychology, 1978, 37, 1-14.

SYMPOSIUM
MENTAL MODELS
OF PHYSICAL PHENOMENA

Generative Analogies as Mental Models

Dedre Gentner

> Subject explaining electric current: If you increase resistance in the circuit, the current slows down. Now that's like a high--cars on a highway where you--if you notice as you close down a lane, you have cars moving along. Okay, as you go down into the thing, the cars move slower through that narrow point.

When people are reasoning about an unfamiliar domain, they often appear to use analogies, as in the above example from a protocol. Analogies are also used in teaching, as in the following excerpt:

> The idea that electricity flows as water does is a good analogy. Picture the wires as pipes carrying water (electrons). Your wall plug is a high-pressure source which you can tap simply by inserting a plug. The plug has two prongs--one to take the flow to the lamp, radio, or air conditioner, the second to conduct the flow back to the wall. A valve (switch) is used to start or stop flow.

If implicit or explicit analogies are an important determinant of the way people think about complex systems, then it becomes crucial to know exactly how such analogies work. This paper considers the psychological role of these analogies in structuring the target domain.

The first question that must be posed is whether such seeming analogical models do in fact strongly affect the person's conceptualization of the target domain (the Generative Analogy hypothesis), or whether they are merely convenient ways of talking about the domain (the Mere Terminology hypothesis). The mere use of terms borrowed from a given domain--as, for example, when electricity is discussed in terms of moving vehicles or flowing water--is not in itself proof that the speaker is conceiving of electricity as deeply analogous to traffic or to water flow.

To demonstrate that an analogy has generative conceptual power, we must show that nontrivial inferences specific to the base occur in the target. These inferences must be such that they cannot be attributed to shallow lexical associations; e.g. it is not enough to find that the person who speaks of electricity as "flowing" also uses terms such as "capacity" or "pressure". Such usage is certainly suggestive of a Generative Analogy, but it could also occur under the Mere Terminology hypothesis.

The goal here is to show that, at least some of the time, the Generative Analogy hypothesis holds: that deep, indirect inferences in the target follow from use of a given base domain as an analogical model. To do this, we must first decide what inferences should follow from use of a given analogy, and then observe whether the analogies people adopt appear to affect the set of inferences they readily make.

The plan of this paper is (1) to propose a structure-mapping theory of analogy that will allow us to predict the set of inferences that should follow from use of a given analogy; (2) contrast two analogical models for the domain and show that they lead to different indirect inferences; (3) to show that people's inferences concerning simple circuits vary according to which of these models they use; and finally (4) to discuss the general issue of analogical models and structure-mapping.

A structure-mapping theory of analogy. The claim here is that analogies select certain aspects of existing knowledge, and that this selected knowledge can be structurally characterized. First, let's consider what an analogy is not. An analogy such as

> (1) An electric circuit with a battery and resistor much like a plumbing system with a reservoir and a constricted section of pipe.

clearly does not convey that all of one's knowledge about the plumbing system should be attributed to the circuit. The inheritance of characteristics is only partial. This might suggest that an analogy is a weak similarity statement, conveying that some but not all of the characteristics of the base system apply to the target system. But this weak characterization fails to capture the distinction between literal similarity and analogical relatedness. Contrast statement (1) with a literal similarity statement like

> (2) A hose is like a pipe.

The literal similarity statement (2) conveys that the pipe and the hose share object attributes--e.g. cylindrical shape--as well as sharing similar relationships with other objects--e.g. CONVEY (hose, water)/CONVEY (pipe, water). Statement (1) also conveys considerable overlap in functional relations: e.g. IMPEDE (resistor, current) /IMPEDE (constriction, water). However, it does not convey overlap of objects and their attributes. The resistor as a separate object need not have any qualities in common with constriction. The analogy, in short, conveys overlap in the system of relations among objects, but no particular overlap in the characteristics of the objects themselves. The literal similarity statement conveys overlap both in relations among the objects and in the attributes of the individual objects.

The analogical models used in science can be characterized as structure-mappings between complex systems. In these analogies, the objects of the known domain, the base domain, are mapped onto the objects of the domain of inquiry, the target domain; the predicates of the base domain--particularly the relations that hold among the nodes--are then applied in the target domain. Structure-mapping analogy asserts that identical operations and relationships hold among nonidentical things. The relational structure is preserved, but not the objects.

Given a particular propositional representation of knowledge we can proceed with an explicit characterization of analogical mapping. A structure-mapping analogy between a target system T and a base system B

This research was supported by the Office of Naval Research and was carried out at Bolt Beranek and Newman and at U.C.S.D. Donald Gentner collaborated on all aspects of this work, but particularly in developing the two analogical models of electricity. I thank Allan Collins, Ken Forbus, Ed Smith and Al Stevens for many insightful comments on this research. Please address correspondence to Dedre Gentner, Bolt Beranek & Newman, 50 Moulton Street, Cambridge, MA 02138.

is an assertion that

1. Given a decomposition of the base and target domains into object nodes $b_1, b_2, \ldots, b_n$ of the base system B and object nodes $t_1, t_2, \ldots, t_m$ of the target system T,
2. The analogical mapping M maps the nodes of B $M: b_i \rightarrow t_i$ into the nodes of T.
3. Then predicates valid in B can be applied in T, using the node substitutions dictated by the mapping:

   $M: \lceil F(b_i, b_j) \rfloor \rightarrow \lceil F(t_i, t_j) \rfloor$

   Further, the probability that the derived proposition is valid in the target T is greater for relational predicates than for attributes, and greater for high-order relations than for lower-order relations. The strength of the analogical predication increases as we move down the list.

   (i) $M: A(b_i) \rightarrow A(t_i)$

   (ii) $M: \lceil F(b_i, b_j) \rfloor \rightarrow \lceil F(t_i, t_j) \rfloor$

   (iii) $M: \lceil G(F_1(b_i, b_j), F_2(b_i, b_j) \rfloor \rightarrow \lceil G(F_1(t_i, t_j), F_2(t_i, t_j) \rfloor$

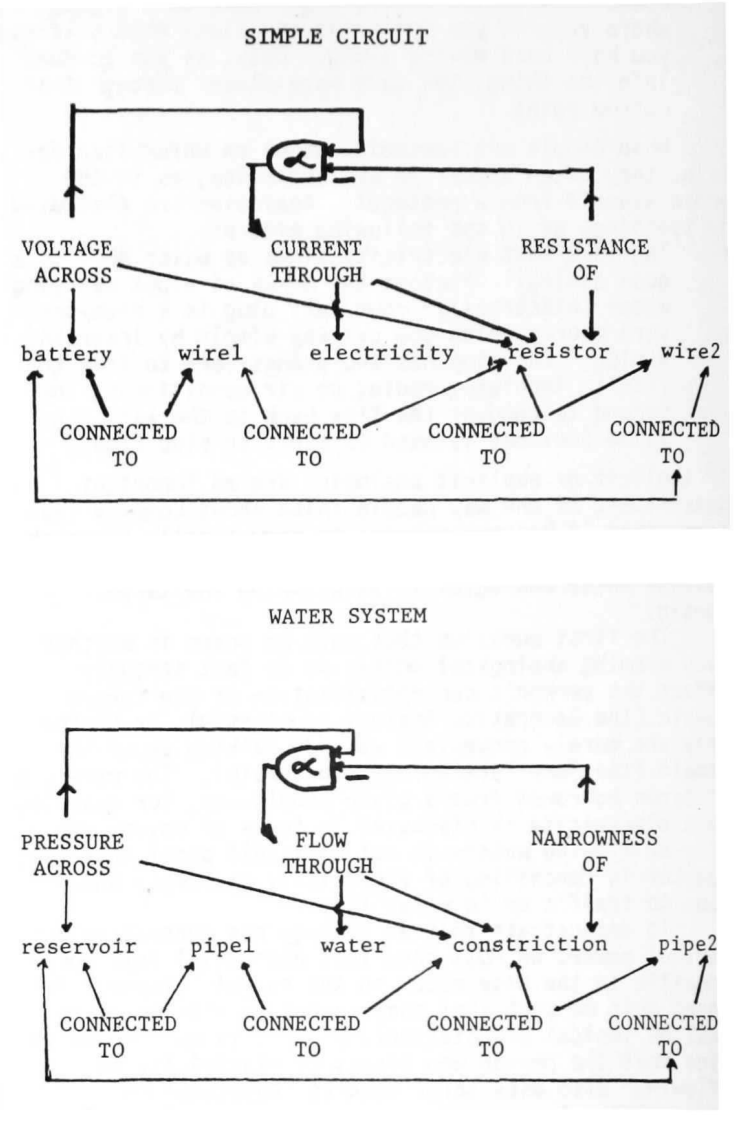
Thus, TRUE $\lceil A(b_i) \rfloor$ does not strongly suggest TRUE $\lceil A(t_i) \rfloor$. Attributional predicates (i) are less likely to carry over than relational predicates (ii); and lower-order relations less likely than higher-order relations (iii).

Two models of simple electric circuits. One common analogy used to teach simple electricity is based on plumbing systems. Figure 1 shows the structure-mapping conveyed by this analogy. The object-nodes of the hydraulic base domain (e.g. the reservoir and constriction) are mapped onto the object-nodes (the battery and resistor) of the circuit. Given this correspondence of nodes, the analogy conveys that the relationships that hold between the objects and object-attributes of the hydraulic system also hold between the nodes of the electric system:for example, that current increases with voltage just as rate of water flow increases with pressure; and that current decreases with resistance just as the rate of water flow decreases with degree of constriction.

A second kind of analogy for electric circuits is based on objects moving through chutes. Current is seen as a moving crowd of small objects: voltage is the forward pressure or pushiness of the objects. Like the plumbing model, the moving-object model provides relations that map usefully into the electrical system: If we imagine a source of pushiness corresponding to the battery, and gates in the chute corresponding to the resistors, then the more pushiness, the higher the rate of aggregate motion; the narrower the gates, the lower the rate of aggregate motion.

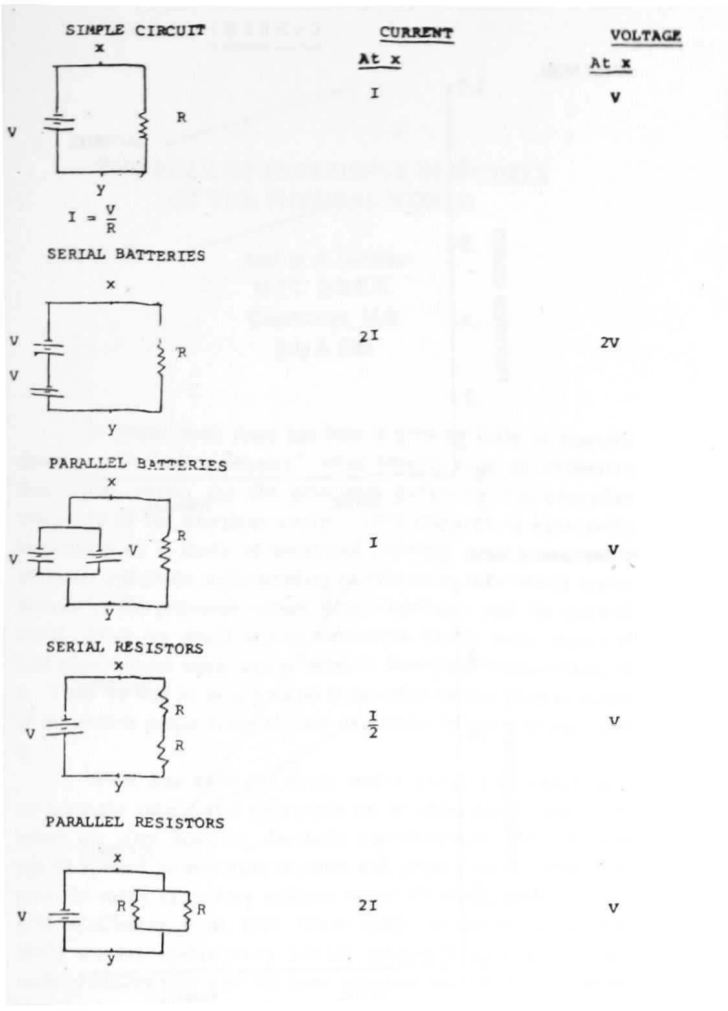
Although these two analogies convey many of the same relations, in some respects they differ in the aptness of the relational match with the target domain, particularly if we consider slightly more complex circuits (see Gentner and Gentner in press.) ing analogy is particularly apt for combinations of batteries, while the moving-object model is superior for combinations of resistors.

Figure 1
Structural representations of water flow and of simple electric circuit, showing structural overlap



Combinational problems. One way to observe deep indirect inferences in the target domain is to ask about different combinations of components. For example, we can ask how the current in a circuit with two resistors in series or in parallel compares with that in a simple one-resistor circuit. The answers are not obvious. The four circuits generated by series and parallel combinations of batteries and resistors are non-transparent. They provide an excellent way to observe true inferences, as opposed to shallow verbal associations. To deduce the current in these four circuits, the person must move beyond the first-stage naive model of circuitry (shown in Figure 1). This first level of insight is that batteries make for more current, resistors make for less current. These rules hold for batteries and resistors in series, but not for parallel combinations, as shown in Figure 2. Parallel batteries give the same current as a single system, not more; while parallel resistors allow more current than in a simple circuit, not less.

Figure 2.
Current and voltage in simple serial and parallel configuration circuits.



Do analogies make a difference. If subjects are really using their analogical models to understand electronics they should draw on their knowledge of combinational relations among the corresponding components in their respective base domains. Even though the complex circuit problems are couched purely in terms of electronics, we should see differences in subjects' predictions depending on which model they use.

For example, here are two sections of a protocol of a subject trying to predict the current in a parallel-resistor circuit. In the first section, she uses a hydraulics model with reservoirs for batteries, which was her initial model, and derives the wrong answer of less current. In the second section, she uses a crowd model suggested to her by the experimenter to derive the correct answer of more current:

HYDRAULICS MODEL WITH RESERVOIR

> We started off as one pipe, but then we split into two. Now does that make any difference? I guess it seems to me that this does make a difference. So what we have here is one pipe, one sort of line coming off and then we let it go like that for a while. We let it split off. We have a different current in the split-off section, and then we bring it back together. That's a whole different thing. That just functions as one big pipe of some obscure description. So you should not get as much current.

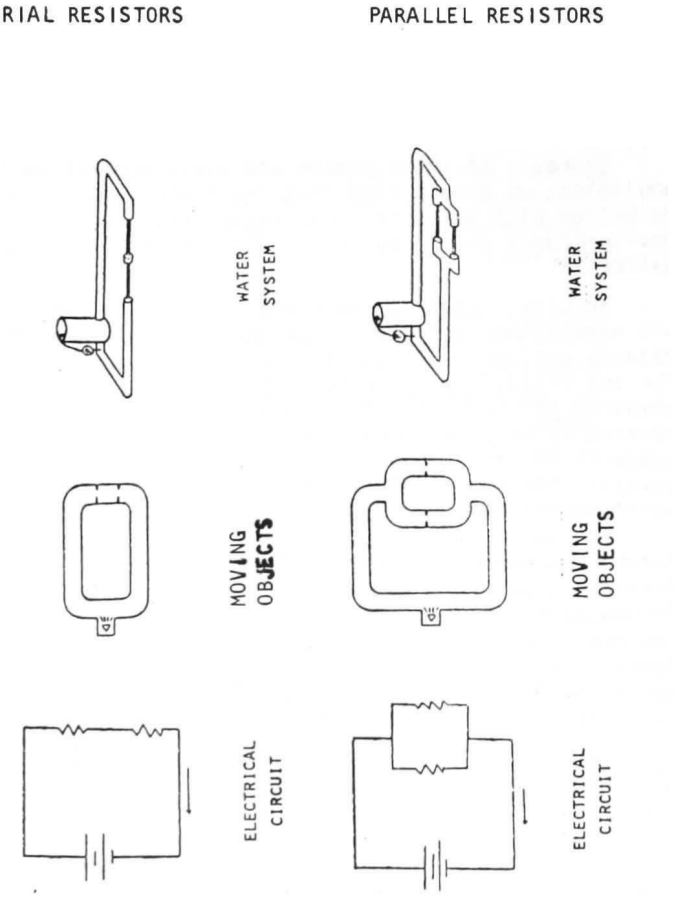
CROWD MODEL

> Again I have all these people coming along here. I have this big area here where people are milling around. I think it is crucial that I separate them though before they get to the gates . . . I can model the two gate system by just putting the two gates right into the arena just like that. So this is one possible model of the two gate system. There are two gates instead of one which seems to imply that the resistance would be half as great if there were only one gate for all those people.

This protocol suggests that models do affect inferences. The following study tests this possibility more on a larger scale. In this study, fairly naive high school and college students were first shown a simple circuit with a battery and a resistor, and then asked to give qualitative solutions for the four combination circuits shown in Figure 2. They were asked to circle whether the current (and voltage) in each of the combination circuits would be greater than, equal to, or less than that of the simple battery-resistor circuit. After they gave their answers for all four combination circuits, they were asked to describe the way they thought about electricity. Then, for each of the four circuit problems, they were asked to circle whether they had thought about flowing fluid, moving objects, or some other way of conceiving of electricity. They were also asked questions about water, to be sure that they understood the base domain.

Figure 3 shows a schematic diagram of parallel and serial resistors in the target and in the two base systems.

Figure 3



Subject who used the fluid model should do well on the battery questions. This is because serial and parallel reservoirs combine in the same manner as serial and parallel batteries, and the combinational distinctions are spatially quite distinct in the water domain. Two equal reservoirs in series (one above the other) give more pressure and hence a greater rate of water flow than a single reservoir. However, since water pressure depends only on height, not on volume, two reservoirs in parallel (side by side) yield only a rate of flow equal to that of a single reservoir. Thus, if the flowing fluid analogy is generative, then subjects with this model (assuming they know the way pressure works in the base) should be able to differentiate serial and parallel configurations of batteries. For resistors, however, the fluid flow model with its constrictions should not, in general, lead to a strong differentiation between serial and parallel resistors. As the protocol above shows, the different between serial and parallel constrictions is fairly opaque in the fluid flow domain; therefore subjects with this model cannot import the correct combinational distinction into the electricity domain. Thus, the prediction is that subjects with the fluid flow model should do better with batteries than with resistors.

For subjects with the moving-objects model, the pattern should be quite different. In this model, configurations of batteries should be relatively difficult to differentiate, since analogies for batteries are hard to find. In contrast, resistors should be better understood. This is because in the moving-objects model, resistors are often seen as gates. Conceiving of resistors as gates should lead to better differentiation between the parallel and serial configurations. If all the objects must pass through two gates one after the other (serial) then the rate of flow should be lower than for just one gate. On the other hand, if the flow splits and moves through two parallel gates, then the rate of flow should be twice the rate for a single gate. Thus subjects using this model should correctly respond that parallel resistors give more current than a single resistor; and serial resistors, less.

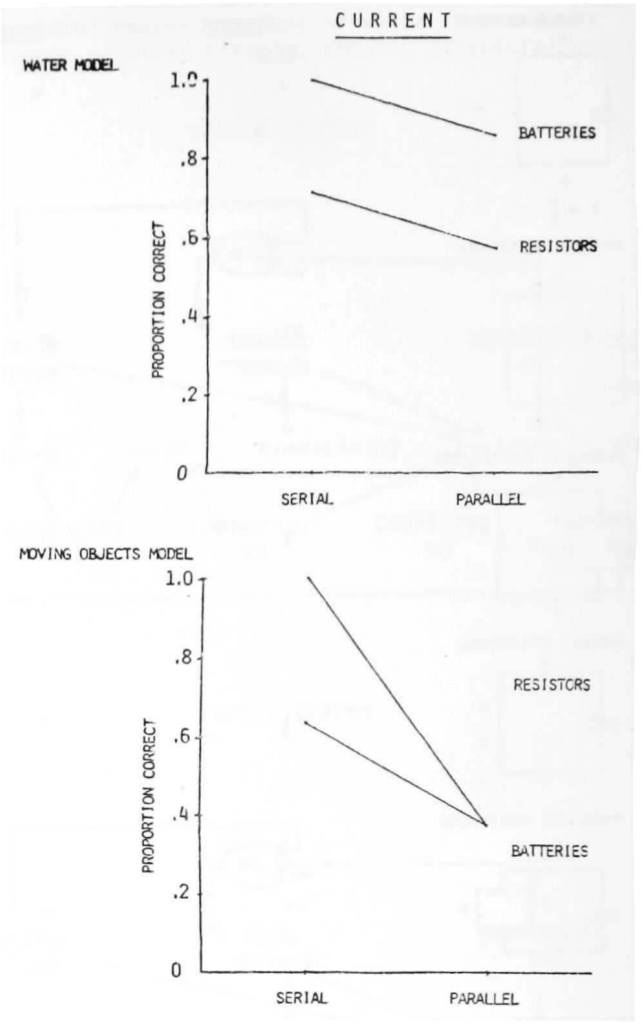
Overall, if these models are truly generative analogies, we should find that the fluid-flow people do better with batteries than resistors, and the moving-object people do better with resistors than with batteries.

Results. Figure 4 shows the results for subjects who used either the fluid-flow analogy or the moving-objects analogy consistently, on all four problems. For the fluid-flow subjects, only those who correctly answered the latter questions about the behavior of reservoirs were included. This was to insure that subjects possessed the requisite knowledge in the base domain. There were nine fluid-flow subjects and seven moving-object subjects.

The patterns of combinational inference are different depending on which model the subject had. As predicted, people who used the fluid-flow model performed better on batteries than on resistors. The reverse is true for the moving-object people. In a Model type x Component type x Topology analysis of variance, the interaction between model type and circuit component is significant; $F(1,13) = 4.53$; $p < .05$.

Conclusions. The results of the study indicate that, for our subjects, the analogies used for electricity were truly generative. Use of different analogies led to systematic differences in the patterns of correct and incorrect inferences in the target domain. Moreover, these combinatorial differences are not easily attributable to shallow verbal associations and communicative patterns. These analogies seem to be truly generative for our subjects; structural relations from the base domain are mapped into the target domain, where they genuinely affect the person's conceptual view of the domain.

Figure 4
Proportion correct on different kinds of circuits for subjects using different models of electricity



The structure-mapping interpretation of the process avoids two extreme positions that often arise in discussions of analogy as explanation: the "vague metaphoricizing" position, which holds that analogy is inherently illogical and unhelpful, and the "appropriate abstractions" position, which emphasizes the fact that analogies convey correct knowledge about the target domain. The structure-mapping view is neutral with respect to whether analogy per se is helpful or harmful. According to the structure-mapping view, the inferences conveyed by a given analogy are not necessarily either correct or incorrect; they are predictable from the predicate structure in the two domains. Relational predicates, particularly those that participate in higher-order systems of relations, are most likely to be mapped; but these may be either correct (as in the case of the moving-object model applied to parallel resistors) or incorrect or indeterminant (as when the moving-object model is applied to batteries). The more we know about the structure of analogy, the better we can design good educational analogies and predict the problem that will occur in use of any given analogy.

References

Gentner, D. and Gentner, D.R. Flowing waters or teeming crowds: Mental models of electric circuits. In D. Gentner & A. Stevens (Eds.) Mental models. Hillsdale, New Jersey: Lawrence Erlbaum Associates in press.

THE ROLE OF EXPERIENCE IN MODELS OF THE PHYSICAL WORLD

Andrea A. diSessa
M.I.T. D.S.R.E.
Cambridge, MA
July 8, 1981

In recent years there has been a growing body of research dealing with "naive physics:" what people, prior to extensive instruction, expect are the principles governing the everyday workings of the physical world. This research is extremely interesting as a study of untutored learning since presumably whatever systematic understanding physics-naive individuals arrive at must be due primarily to their direct experience with the physical world. Here we would like to summarize briefly some results of that experimental work and a tentative theoretical interpretation of it. Then we will be in a position to speculate on the general nature of the models people make of their experience in order to cope with it.

What has emerged from naive physics research is a surprisingly robust and systematic set of ideas about mechanics which are often, however, decidedly non-Newtonian. Studies from age 10 upward to university students and physics-naive adults have given in many cases very uniform results (Viennot, 1979; Clement 1979; McCloskey, *et al*, 1980; White, 1981). In one study (diSessa 1981) we let elementary school students and university undergraduates play with the same computer simulation of an object obeying Newton's Laws. Despite many more years experience, a year of high school physics and a university level course on mechanics one could see a clear overlap in the set of strategies the university students used as compared to the elementary students. Moreover, many of these common strategies were not neutral, not merely based on other-than-textbook analyses, but were overtly non-Newtonian. The university students often had difficulty applying the simplest classroom concepts to the simulation, even when asked. One of the prominent expectations most students showed was that force acts by directly producing motion in the direction of the force rather than by combining with previous motion. Thus these students sided with Aristotle against Newton.

Besides the obvious pedagogical problems, data such as this pose in a very direct form the fundamental question of what one learns from experience. How can it be that people come to a robust non-Newtonian understanding, even resisting instruction, of a world with which they deal everyday, a world governed by Newtonian principles?

No one expects the answers to such a question to be simple. But some analysis of a set of naive conceptions (diSessa, to appear) suggests that something like the following mechanism may play an important role. It is a mechanism concerning incremental learning based on experience. The basic idea is that among all the experiences one has regarding a class of phenomena, for example pushing and pulling things around, a few are selected to stand as prototypical for the class and are systematically used in both explanations and predictions. It is not, of course, a literal recall of an observed phenomenon which serves this purpose but what the person establishes as a conventional interpretation of the phenomenon in terms which that person already understands. I call these paradigmatic interpretations of experience "phenomenological primitives." In the case of the Aristotelian expectation of motion always in the direction of force, one might hypothesize that the common event of pushing an object from rest serves as an important prototype. Thus the phenomenological primitive here is the "theory" that things simply move in the direction you push them, and previous momentum is generally ignored since it plays no role in the prototype. I suspect that the interpretation of pushing from rest is based on prior, common-sense notions of agency and causality which one finds associated with naive conceptions of force in many ways. Indeed, what is more surprising, one can make a rather strong case that these same ideas had a great impact on the historical development of the science of mechanics as well. (See diSessa, 1980.)

Though the notion of phenomenological primitive might help account for the origins of false "theories" like the Aristotelian expectation, we must look to the knowledge system in which these structures operate in order to account for their long term stability. Here we can offer only the briefest suggestion as to the character of this system, again by example. The example involves a counter-example to the Aristotelian expectation. Imagine a ten-ton truck hurtling down the highway and a small push on its side. Who could believe the truck will move in the direction of the push? Indeed, no one does! When subjects are prompted in such a way, the Aristotelian intuition is not even considered, or, if it is, excuses for its inapplicability are found: "The force (sic) of the truck is too big; it's overcoming the sideways push." But either way, (through selective cueing or maintaining excuses as part of the knowledge base) the Aristotelian expectation is isolated and kept safe from refutation.

What generalizations can we draw from this story of naive physics about the models people spontaneously make? The first is that these models are robust, that naive ideas and the interpretations of experience one makes with them can have a powerful effect on long term understanding or misunderstanding. All the evidence points to the conclusion that naive physics plays a large role in learning textbook physics. To draw a caricature, it is almost as if one is trying to teach physics to a stable cognitive system which already knows a different physics.

The second generality is the importance of knowing the particular pre-existing notions. What one carries away from an experience is one's interpretation of it. Such a truism only warrants attention when we remind ourselves of how little we know about the naive vocabulary and about how one builds deeper understanding out of it. The surprise of a Newtonian world teaching Aristotelian physics, however, should serve as a reminder.

The third generality concerns the fragmentation exhibited by spontaneous models. One can find people both believing and

disbelieving their Aristotelian expectations depending on circumstances. This apparent incoherence is puzzling. After all, what other than a coherent system could exhibit such robustness. In fact, it is in this area, coherence, that I believe our own naive ideas about knowledge systems most need refining.

References

Clement, J. Common preconceptions and misconceptions as an important source of difficulty in physics courses, Amherst, MA: University of Massachusetts, Cognitive Development Project Working Paper, July, 1979.

diSessa, A. Unlearning Aristotelian physics: a study of knowledge-based learning, *Cognitive Science*, to appear 1981.

diSessa, A. Phenomenology and the evolution of intuition, in *Mental Models*, Gentner, D. and Stevens, A. (Eds.), Lawrence Erlbaum, to appear.

diSessa, A. Momentum flow as an alternative perspective in elementary mechanics, (Working Paper •4), Cambridge MA: MIT DSRE, 1980.

McCloskey, M., Caramazza, A. and Green, B. Curvilinear motion in the absence of external forces: naive beliefs about the motion of objects, *Science* Vol. 210, 5, December, 1980.

Viennot, L. *La raisonnement spontane en dynamique elementaire*, Paris: Hermann, 1979.

White, B. Designing computer games to enhance learning, (Technical Report) Cambridge, MA:. MIT AI Laboratory, 1981

THE FORM AND FUNCTION OF MENTAL MODELS

P. N. Johnson-Laird

Centre for Research on Perception and Cognition
Laboratory of Experimental Psychology
University of Sussex
Brighton BN1 9QG England

You are lost in the maze at Hampton Court Palace. You come to a turning and for a moment you are not sure which way to go. You recognize that you have been at this point before, and, in your imagination, you turn right, proceed down an alley, and are then confronted by a dead end. And so this time around, you decide to turn left. What you did was to reconstruct a route through the maze on the basis of a mental model of it. You may hardly have experienced any imagery at all; or you may have had a succession of vivid images like a snippet from an imaginary movie that culminates in a leafy cul-de-sac. In either case, there was nothing verbal about your reasoning: you navigated your way through your model of the maze much as a rat in a psychological laboratory might have done (O'Keefe and Nadel, 1978). Yet, there is another method that you could use to make your decision. You recall instead that the way to get out of the maze is to keep turning left at every available opportunity, and, since you are presented with such an opportunity, you accordingly decide to turn left. This method makes use of a mental representation of verbal propositions.

The two alternatives illustrate the contrast between exploiting a mental model (perhaps with accompanying imagery) and making use of a propositional representation. My aim in this paper is to show that the contrast is real -- that there are both forms of mental representation -- and to offer an account of the purpose that they serve. Indeed, if there are mental models, then the two most important questions about them are: what form do they take? what function do they serve? I will try to answer both questions.

Direct empirical evidence for the contrast between propositional representations and mental models comes from a series of experiments that Kannan Mani and I have carried out (Mani and Johnson-Laird, in press). In the most recent of our studies, the subjects heard a verbal description of a spatial layout, such as:

The spoon is to the left of the knife
The plate is to the right of the knife
The fork is in front of the spoon
The cup is in front of the knife.

They were then shown a diagram, such as:

spoon knife plate

fork cup

and they had to decide whether or not the diagram was consistent with the description. (If you think of the diagram as depicting the arrangement of the objects on a table top, then obviously it is consistent with the description.) Half the descriptions that the subjects received were determinate like the example above, and the other half were indeterminate. The indeterminate descriptions were constructed merely by changing the last word in the second sentence:

The spoon is to the left of the knife
The plate is to the right of the spoon
The fork is in front of the spoon
The cup is in front of the knife.

This description is consistent with two radically different diagrams:

(1)			(2)		
spoon	knife	plate	spoon	plate	knife
fork	cup		fork		cup

The materials were counterbalanced so that for each set of five objects, a subject received either the determinate or else the indeterminate description. After the subjects had judged a series of eight descriptions and diagrams, they were given an unexpected test of their memory for the descriptions. On each trial, they had to rank four alternatives in terms of their resemblance to the original description: the original description, an inferrable description, and two 'foils' with a different meaning. The inferrable description for the example above contained the sentence:

The fork is to the left of the cup

in place of the sentence interrelating the spoon and the knife. The description can therefore be inferred from the layout corresponding to the original description in the case of both the determinate and the indeterminate descriptions.

The subjects remembered the layouts of the determinate descriptions very much better than those of the indeterminate descriptions. The percentages of trials on which they ranked the original and the inferrable descriptions prior to the foils was 88% for the determinate descriptions, but only 58% for the indeterminate descriptions. All twenty of the subjects conformed to the trend, and there was no effect of whether or not a diagram had been consistent with a description. However, the percentages of trials on which the original description was ranked higher than the inferrable description was 68% for the determinate descriptions, but 88% for the indeterminate descriptions. This difference was highly reliable, too.

Evidently, subjects tend to remember the layout of determinate descriptions better than that of

indeterminate descriptions, but they tend to remember verbatim detail of indeterminate descriptions better than that of determinate descriptions. This 'cross-over' effect is impossible to explain without postulating at least two sorts of mental representation. A plausible account of the results is indeed that subjects construct a mental model of the determinate descriptions but abandon such a representation in favour of a superficial propositional one as soon as they encounter an indeterminacy in a description. Mental models are easier to remember than propositional representations, perhaps because they are more structured and elaborated (cf. Craik and Tulving, 1975) and require a greater amount of processing to construct (cf. Johnson-Laird and Bethell-Fox, 1978). But, models encode little or nothing of the linguistic form of the sentences on which they are based, and subjects accordingly confuse inferrable descriptions with the originals. Propositional representations are relatively hard to remember, but they do encode the linguistic form of sentences. Hence, when they are remembered, the subjects are likely to make a better than chance recognition of verbatim content.

It is natural to suppose that propositional representations are produced as part of the normal process of comprehending discourse (cf. Kintsch, 1974; Fodor, Fodor, and Garrett, 1975), but they can also serve the useful purpose of providing an economical representation of radically indeterminate discourse. The function of mental models is profoundly semantic: a propositional representation is true or false with respect to a mental model of the world. This relation is established by mapping propositional representations onto mental models, and I have argued elsewhere for a procedural semantics that carries out this task (see, e.g. Johnson-Laird, 1980, for a description of a program that builds up spatial arrays from verbal descriptions). Truth or falsity with respect to reality ultimately depends on the construction of mental models on the basis of perceptual experience.

There is one other crucial function served by mental models. The fundamental semantic principle of truth is that an assertion is true provided that there is no counterexample to it. The assertion, "Socrates is dead," has only one possible counterexample, namely, that Socrates is not dead; the assertion, "All men are mortal," has a large number of potential counterexamples. Likewise, given the truth of a set of premises, a conclusion is necessarily true only if there is no counterexample to it, that is, no way of interpreting the premises that renders the conclusion false. If human beings have grasped this principle, then they can reason validly without possessing any mental logic, rules of inference, or inferential schemata. It is a straightforward matter to write computer programs that make inferences without recourse to rules of inference: they construct models of the premises, draw a putative conclusion on the basis of a simple heuristic, and then search for counterexamples to the conclusion. On the previous occasion that I presented this idea (Johnson-Laird, 1980), it was viewed as on a par with the Pelagian heresy in some quarters. Yet cognitive scientists should be prepared to accept that the doctrine of mental logic may be just as mistaken as the idea of original sin. Abandoning the doctrine certainly solves the otherwise intractable mystery of how children could acquire logic without being able to reason validly. The thesis that logic is innate is the only plausible solution but it has no more explanatory value or empirical content than an appeal to divine intervention. If there is no mental logic, then the question of of its origins does not even arise. The theory of mental models also reveals the major cause of inferential error: the greater the number of mental models that have to be constructed in order to make a valid deduction, the greater load on working memory, and the more likely an error is to be made. My colleagues and I have checked this prediction in a variety of inferential tasks. Table 1 presents the relevant data from four of our experiments in which the subjects had to draw their own conclusions from syllogistic premises: it gives the percentages of valid conclusions that were drawn depending on the number of mental models that had to be

Table 1: The percentages of correct valid conclusions drawn from syllogistic premises in four experiments. The percentages are shown as a function of the number of mental models that have to be constructed in order to draw a valid conclusion.

	One model problems	Two model problems	Three model problems
Experiment 1	92	46	28
Experiment 2	80	20	9
Experiment 3	62	20	3
Experiment 4	58	0	0

constructed. In Experiment 1, 20 students at Teachers College, Columbia University, were asked to state what followed from premises in each of the 64 logically distinct varieties (see Johnson-Laird and Steedman, 1978). Experiment 2 was a replication with 20 students at Milan University, and Experiment 3 was a further replication in which 20 Italian subjects were given just 10 seconds in which to make each of their responses. These experiments were carried out in collaboration with Bruno Bara. Finally, in Experiment 4, which Debbie Bull and I designed, 19 children between 11 and 12 years of age were asked to draw conclusions from 20 out of the 64 possible pairs of syllogistic premises. The trend in each experiment was remarkable: not a single subject that we have tested has ever failed to perform best on those syllogisms that require only a single model to be constructed. It is difficult to resist the conclusion that inferential ability is based on the manipulation of mental models.

Let me finish with one final conjectural flourish. The psychological core of understanding any phenomenon consists in your having a 'working model' of it in your mind. If you understand inflation, a mathematical proof, the way a computer works, DNA or a divorce, then you have a mental representation of it that serves as a model in much the same way as, say, a clock functions as a model of the solar system. Like a clock, a mental model need not be wholly accurate to be useful, which is just as well because, of course, there are no complete models of any empirical phenomena. If a television set is mentally represented as containing a beam of electrons that are magnetically deflected across the screen, then this component of the model serves an explanatory function. It accounts, for example, for the distortion of the picture that occurs when a magnet is held near to the screen. Other components of the model may serve no such function. One might imagine, say, each electron as deflected by the magnetic field much as a ball-bearing is diverted from its course by a magnet, but without having any representation of the nature of magnetism: the 'picture' is just a picture, which simulates reality rather than models its underlying principles. At least one other component of every dynamic model is

neither modelled nor simulated. This element is time. Time is not represented in a dynamic model, but rather the model unwinds in real time in much the same way as do the events that are modelled, though perhaps at a different rate. In models that are not dynamic, of course, time can be represented by a spatial axis. What one should expect in examining the growth of expertise in a particular domain is the gradual transition from mere propositional principles to a fully articulated mental model, and the gradual replacement of simulated elements by their modelled counterparts.

REFERENCES

Craik, F.I.M., & Tulving, E. Depth of processing and the retention of words in episodic memory. Journal of Experimental Psychology: General, 1975, 104, 268-294.

Fodor, J.D., Fodor, J.A., & Garrett, M.F. The psychological unreality of semantic representations. Linguistic Inquiry, 1975, 4, 515-531.

Johnson-Laird, P.N. Mental models in cognitive science. Cognitive Science, 1980, 4, 71-115.

Johnson-Laird, P.N., & Bethell-Fox, C.E. Memory for questions and amount of processing. Memory and Cognition, 1978, 6, 496-501.

Johnson-Laird, P.N., & Steedman, M.J. The psychology of syllogisms. Cognitive Psychology, 1978, 10, 64-99.

Kintsch, W. The representation of meaning in memory. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Mani, K., & Johnson-Laird, P.N. The mental representation of spatial descriptions. Memory and Cognition, In press.

O'Keefe, J., & Nadel, L. The hippocampus as a cognitive map. London: Oxford University Press, 1978.

Learning Through Growth of Skill in Mental Modeling[1]

Jill H. Larkin
Herbert A. Simon
Carnegie-Mellon University

The mental models of a skilled scientist are often different from those of an untrained person. For example, in thinking about the interaction of physical objects, the untrained person seems largely restricted to envisioning objects in a sequence of motions. The entities in these "naive" or *immediate* mental models correspond directly to objects in the real world. The inferential rules that control the running of these models correspond roughly to rules reflecting how events unfold in real time. While such mental models are perfectly adequate for getting around in everyday situations, they are sometimes dramatically wrong (Green, McCloskey, and Caramazza, 1980, Clement, 1981) and they certainly seem less effective in solving scientific problems than are the more extended representations of the scientist.

The mental models of a person with training in physics are not limited to entities and inferential rules based directly on experience. Instead, these models can and do include entities that have technical meanings defined only by the scientific discipline, and that are related by special inferential rules again defined in the discipline. For example, a person with training in physics is not restricted to considering perceivable objects like cats and coffee cups, but may also represent situations in terms of technical entities like forces or pressure drops. Similarly, capacity to make inferences about a situation need not parallel imagined development of the situation in time, but instead may reflect special constraint laws of physics, e.g., that the momentum of an isolated system must remain constant. These ideas are discussed more fully in (Larkin, 1981a) and in the discussion of physical intuition in (Simon and Simon, 1978).

In this paper we consider how an individual might develop the ability to re-represent situations in terms of scientific entities. Presumably this development is one goal of science instruction. We shall present preliminary results from an experimental and theoretical case study in such development. Subjects with backgrounds in physics studied sections taken from a physics textbook that described material (fluid statics) they had not previously encountered, and then used this material in efforts to solve problems. In a coordinated theoretical effort we are developing a computer-implemented model of learning from text that is capable of using declarative statements of facts (in this case, relations of physics) both to "understand" the derivation of new results and to apply these results in solving problems.

1. The ABLE system

The system, called ABLE, is a descendant of a system that learned through practice to apply principles of mechanics, and that accounted for strategy differences between skilled and less skilled individuals (Larkin, 1981b).

ABLE is a *production system* written in the current implementation of OPS, a LISP-based efficient production-system language (Forgy, 1980, Forgy, 1979). Thus ABLE has a *working memory* composed of passive elements of knowledge that are acted on by a large *production memory* composed of elements of procedural knowledge encoded as condition-action pairs. When the conditions of a particular production are found to match some of the contents of working memory, this match cues the execution of the corresponding actions which then act to modify working memory. Then the conditions of some new production are found to match, and the cycles continue. Production systems have a continued history of fruitfulness in psychological modeling (Newell and Simon, 1972, Newell, 1973, McDermott, 1978) but the major feature used in ABLE is the easy modeling of learning. Each production is an independent piece of knowledge, and the circumstances under which it applies are determined only by the contents of its own conditions. Thus the addition of knowledge (learning) is modeled simply by the addition of new productions.

To explain the working of ABLE we consider its application to solving part of the problem given in Table 1 and Figure 1a, and presented as a worked example in Halliday & Resnick (1970). The first paragraph of the example states the problem. We currently give ABLE a good understanding of this paragraph, i.e., a good immediate representation of the problem. It is coded as a set of related declarative elements in working memory, indicated by the graph structure in Figure 1b.

1.1. Encoding of Principles

After achieving this immediate representation of the situation, how does a solver make scientific inferences of the kind illustrated by the textbook solution given in Table 1? In other words how is a scientific mental model run?

Such inferences must be based on scientific principles that are in some sense "known" to the solver. "Known" might initially mean that the appropriate textbook page is available for inspection. We discuss later the growth of other kinds of knowing. Thus we provide ABLE with knowledge of relevant principles in the form illustrated in Figure 2. Like all principles and definitions in ABLE, it includes a symbolic *statement* of the principle $\Delta p \quad \rho gh$ together with a *setting* to which the principle applies and in terms of which of the symbols in the statement are defined. Here the setting includes a portion of liquid with density ρ, two points in that liquid separated by a height

Table 1: Worked example from a textbook (Halliday and Resnick, 1970) showing the application of relations of fluid statics to relate densities of liquids in a U-shaped tube.

A U-tube is partly filled with water. Another liquid, which does not mix with water, is poured into one side until it stands a distance d above the water level on the other side, which has meanwhile risen a distance l (Fig. 1a). Find the density of the liquid relative to that of water.

In Fig. 1 points C are at the same pressure[1] Hence, the pressure drop from C to each surface is the same[2], for each surface is at atmospheric pressure[3]

The pressure drop on the water side is $\rho_w 2l^4$, where the $2l^5$ comes from the fact that the water column has risen a distance l^6 on one side and fallen a distance l on the other side, from its initial position. The pressure drop on the other side is $\rho g(d + 2l)^7$, where ρ is the density of the unknown liquid. Hence,

$$\rho_w g2l = \rho g(d + 2l)^8$$

and

$$\rho/\rho_w \quad 2l/(2l + d)^9.$$

The ratio of the density of substance to the density of water is called the *relative density* (or the *specific gravity*) of that substance.

[1-9] These numbers label inferences for reference later in the text.

h. The "gravitational acceleration" $g = 9.8 m/s^2$ is not specified but assumed to be known outside the context of this principle.

This knowledge of a principle is encoded as a passive link-node structure involving no knowledge of how or when to apply the principle. In this sense it is *declarative* knowledge, although clearly

[1]This work was supported by NIE-NSF grant number 1-55862, by NSF grant number 1 55035 and by the Defense Advanced Research Projects Agency (DOD), ARPA Order No. 3597 monitored by the Air Force Avionics Laboratory under Contract F33615-78 C 1151. The authors acknowledge the important contributions of Susan Cotten in the collection, transcription and coding of the protocol data.

Figure 1: (a) Diagram provided by the textbook for the example in Table 1. (b) Annotated "diagram" (immediate representation) provided for ABLE in starting to work the example.

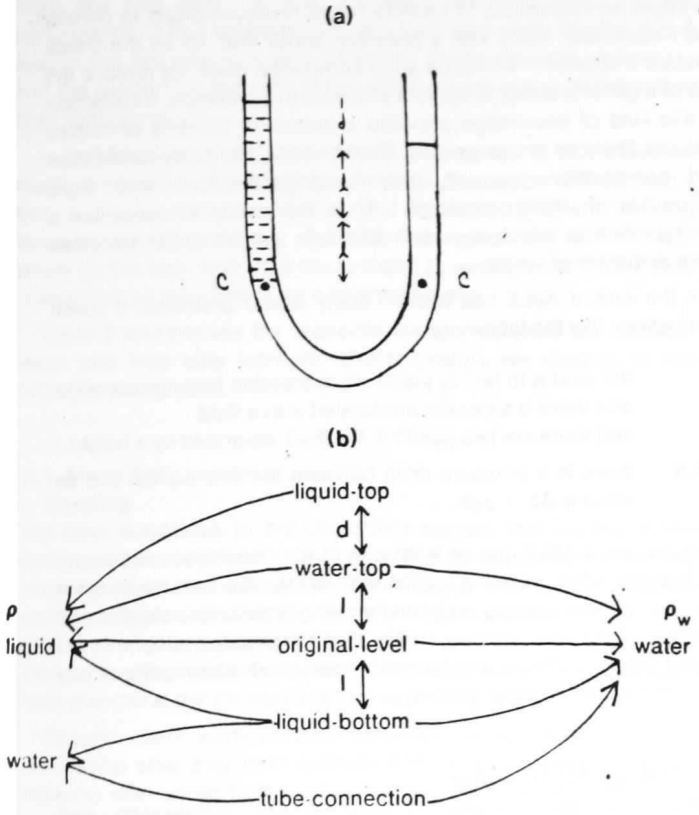
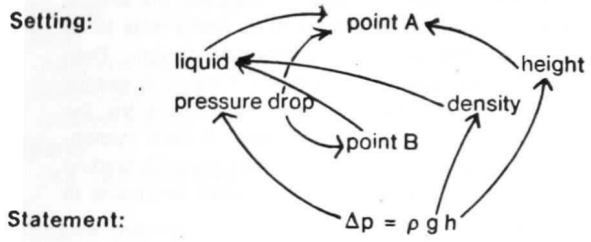


Figure 2: Graph structure representation of the principle $\Delta p = \rho gh$.



it goes beyond minimal propositional encoding of the phrases that may have been used to describe it in the textbook. To illustrate how ABLE uses this knowledge of principles, we consider its application to develop the inference labeled 4 in Table 1, that is to infer that the pressure drop from C to A on the water side of the tube is $\rho g2l$.

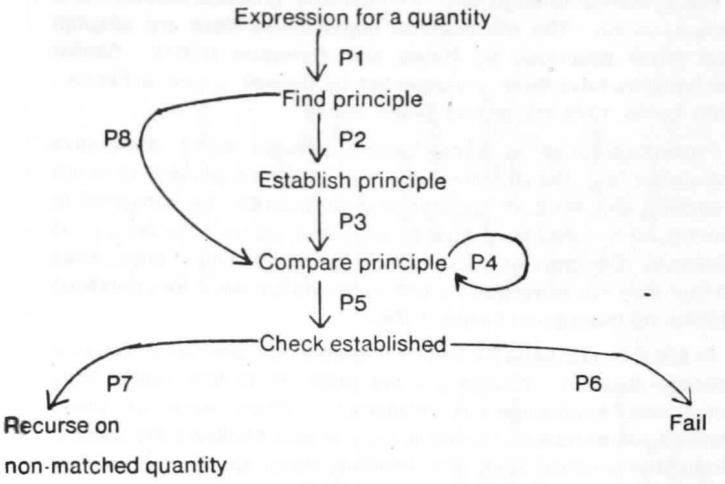
1.2. Interpretive Use of Principles

ABLE applies declarative knowledge through general procedural knowledge that first matches the setting of the principle to the setting of the problem, and then uses the statement of the principle (interpreted in the setting of the problem) to make inferences.

The control structure of ABLE, shown in Figure 3, is based on that proposed by Neves and Anderson (1981) for the analogous task of supplying reasons for the statements in a geometry proof. The following paragraphs provide English statements of the productions that control the shift of goals, Figure 3 indicates these productions by labeled arrows linking goal statements.

ABLE starts by using the knowledge in production P1 that if the

Figure 3: Goals used in ABLE with labeled arrows indicating productions that change goals.



goal is to find an expression for a quantity in a problem setting, then one should search for a principle that can provide further information about that quantity. The initial search can be based on a variety of criteria (cf. Simon & Simon (1978), Larkin 1981b) . ABLE currently uses a very rough process, embodied in P2, that to be considered a principle must involve a quantity of the type currently desired (here pressure drop).

ABLE then applies the knowledge in production P3 to set the goal of comparing elements of the problem situation with elements of the principle, including its statement and setting. Production P4 contains knowledge of how to trace and compare the graph structures representing the current problem situation (e.g., Figure 1b) and the principle situation (Figure 2). When all possible correspondences between the two graph structures have been matched, production P5 sets as a goal to check whether all parts of the principle have been matched. If they have, the goals are successively marked as succeeded. If this is not the case, production P6 recognizes that part of the situation crucial to the applicability of the principle has not been satisfied. The goals are then successively marked as failed, and ABLE ultimately must seek a different principle. If, however, the only part of the principle situation that does not have a correspondence in the problem situation is a particular theoretical entity (quantity) then production P7 recognizes that if this correspondence could be established then the principle would succeed. Thus in our example, if ABLE had not already established the correspondence $h = 2l$ (as the text solution has not -- see Table 1), then ABLE would set as a subgoal to establish an expression for the height h, beginning work again with production P1 in Figure 3.

Much of ABLE's work involves establishing a detailed match between a principle setting and a subset of the problem setting. In the single inference discussed above, of the total of 9 cycles of production execution, 5 were concerned with matching settings. This costly and compulsive matching seems, however, to be necessary for good problem solving. For example, in the current problem there are several densities, pressures, and heights. Without careful matching between settings, the solver can easily "infer" relations between quantities that in fact have no connection.

1.3. Reducing the Costs of Interpretive Matching

Because matching a principle to a setting is costly, it is crucial that a solver develop good search procedures for locating principles likely to produce useful inferences. The primitive search algorithm embodied in production P1 (pick a principle involving the kind of quantity you're trying to solve for) is certainly not good enough. The

following paragraphs describe ABLE's ability to develop better search mechanisms.

ABLE learns through two mechanisms *proceduralization* and *generalization*. The mechanisms implemented here are adapted from those described by Neves and Anderson (1981). Similar mechanisms have been implemented by (Newell, Shaw, & Simon , 1960, Lewis, 1978, Hayes and Simon, 1974).

Proceduralization is a mechanism through which declarative knowledge (e.g., the statement of a principle and a situation to which it applies) can, through application in an example, be converted to procedural knowledge of how to apply that principle to analogous situations. Composition is the collapsing or compiling of procedures so that they run more quickly and with reduced need for conscious monitoring (Hayes and Simon, 1974).

In ABLE productions P1 and P4 (Figure 3) contain the capacity for proceduralization. (These are the productions that use directly declarative knowledge of relations.) When each of these productions executes, it builds a copy of itself involving the specific declarative entities from the relation being applied. Thus for example, when P4 applies to the principle $\Delta p \quad \rho gh$, it may apply in the form:

IF If the goal is to compare a principle to the current setting
and there are corresponding theoretical entities in the principle and problem settings
and these entities refer to two physical entities of the same type

THEN mark the two physical entities as corresponding.

This production builds a copy of itself of the form:

IF the goal is to compare the principle $\Delta p = \rho gh$ to the current setting
and the density of a liquid in the problem setting corresponds to the density in the principle setting

THEN mark this liquid in the problem as corresponding to the liquid in the principle $\Delta p = \rho gh$.

This new production is part of the procedural knowledge needed to apply the principle $\Delta p = \rho gh$.

As Neves and Anderson (1981) point out, these proceduralized productions are almost always shorter (contain fewer conditions) than the original general productions that built them. Thus they may immediately provide the advantage of reducing working-memory load. However, their main importance is that they are the ingredients for building efficient productions that recognize useful configurations in a problem and relate these configurations to potentially useful principles.

The second mechanism of learning is composition. If two proceduralized productions (any of those built by P1 and P4) execute in sequence, they are combined to form a single production that does the work of both. This is done first by collecting the condition and action elements from both and deleting repetitions. (Further details given by Neves and Anderson (1981).) For example, the proceduralized production above can be composed with a production built by P1 in Figure 3 to form the following:

IF the goal is to find an expression for Δp of a fluid
and there is a density for the fluid

THEN set the goal to compare to the current situation the principle $\Delta p = \rho gh$
with the correspondence between the pressure drops and the densities already established.

Productions of this form, indicated by P8 in Figure 3 short-circuit the primitive search algorithm of P1. They do not, however, short circuit all the interpretive matching between the principle and problem settings. In the current case, ABLE must still match lines and points and heights.. In human language the knowledge in a production like P8 might be expressed, "I have to relate pressure-drops to density and I remember there was a principle about that, so let me check whether it applies." Although such knowledge does not replace the use of a general ability to apply a principle in a situation, a collection of this kind of knowledge provides a means of locating principles that are likely to prove useful. Furthermore, as proceduralization and composition proceed, they build productions with more information in their conditions. Thus the ability to recognize a configuration in which a certain principle will be useful becomes more and more accurate.

In the limit, if ABLE has worked many similar problems, it builds productions like the following:

IF the goal is to find or justify an expression for pressure drop
and there is a density associated with a fluid
and there are two points in the fluid separated by a height

THEN there is a pressure drop between the two points, and its value is $\Delta p = \rho gh$.

At this point ABLE has at least one of the capabilities necessary for building what we have called a scientific representation for a problem. On encoding a problem involving appropriate heights and densities, ABLE immediately knows that the problem also involves related pressure drops, and can use these entities as readily as any in the immediate representations.

1.4. Settings for Learning

When can the learning involved in proceduralization and composition occur? Clearly solving problems is one such setting. However, as we shall see in our discussion of human learners, it seems likely that this learning also takes place during study of text material. The following is one mechanism through which this might occur.

Consider the textbook example in Table 1. Suppose the learner considers the various inferences (labeled 1-9) as statements to be understood or verified on the basis of previous knowledge. Then understanding the sentence involving inference 4 involves exactly the process discussed above, justifying the expression for the pressure drops by using the principle $\Delta p \quad \rho gh$. If such reason-giving is part of active study of scientific text, then through reading the learner should acquire some ability to recognize situations to which principles will apply.

These comments are not limited to the study of text examples. New principles themselves are presented in much the same manner. A setting is described, and a sequence of inferences are stated, ultimately leading to a statement of a new principle.

2. Human Learners

In previous work (Simon and Simon, 1978. Larkin, McDermott, Simon, and Simon, 1980a, Larkin, 1981b), we have compared the problem solving of true experts, individuals with extensive professional experience in physics, with that of novices, individuals whose experience is limited to the equivalent of less than one college level course. The performance of the expert subjects would correspond here to the ABLE model after all proceduralization and composition had occurred. This very ABLE model is essentially equivalent in performance to the models of expert subjects described in earlier papers. Proceduralization and composition have produced a collection of sensitive productions that can recognize a configuration of knowledge in a problem situation, and make an immediate inference based on an appropriate principle.

The human solvers considered here are all novices with respect to the physics material, they knew varying amounts of general physics.

but none had previously studied fluid statics. The question is then to what extent do human learners, confronted with a novel section of physics text and associated problems, perform like the ABLE system? The answer is that in interesting ways human learners are more and less able. In this discussion we shall focus on the performance of four subjects, two who look very much like the ABLE system and two who look very different. These data are preliminary and the support for features of the ABLE model is suggestive rather than conclusive.

In individual sessions each subject was asked to talk aloud as much as possible while reading a six-page discussion of fluid statics from a physics textbook (Halliday and Resnick, 1970) and working three associated problems. Subjects were given one and one-half hours for the task, and were encouraged to work just as if they were completing an assignment for a science course.

Table 2 summarizes the characteristics of subjects we consider more and less able learners, characteristics we discuss in the following paragraphs.

2.1. More Able Learners

Reading

We have suggested in the preceding section that careful active reading of text may be an important setting for acquiring the partially proceduralized and composed productions that aid in locating useful principles. Indeed the two more able subjects used a great deal of effort in processing the text. First, both subjects began their work by reading the text completely from beginning to end, although both glanced at the problems before beginning to read.

Second, these learners show consistent evidence that they are processing what they read carefully and conscientiously. The two subjects interrupted their reading 53 and 70 times respectively to make comments on what they were reading. Most of these comments (75%, 81%) suggest that the subject is relating what is being read to previous knowledge ("Ok, intuition would tell you that"; "...which is analogous to just the weight of something in mechanics."), or expressions of understanding (e.g., "Ok, that's easy enough").

Table 2: Characteristics of more and less able learners.

More Able	Less Able
Reading:	
Aloud, many comments	Often silently, usually few comments
Reading precedes problem solving	Problem solving before all text read
Slowly (e.g.,19 min, 20 min)	Rapidly (e.g.,7 min, 5 min)
Problem solving:	
Correct or factor of 2 error	Other errors
No search	Search common
Order of principles like ABLE	Order of principles means-ends

Problem Solving

We consider here performance on the following problem, which is very analogous to the worked example in the text (Table 1).

> A simple U-tube contains mercury. When 13.6 cm of water is poured into the right arm, how high does the mercury rise in the left arm from its initial position?

The text provides the densities of water ($1.0 \times 10^3 kg/m^3$) and mercury ($1.36 \times 10^4 kg/m^3$).

The two subjects solved this problem without any search through the text. They readily recalled the relevance of the U-tube example, found the right page in the text, and based their solutions on inferences made in that example. Thus before beginning the problem solution, these subjects had some internalized knowledge of useful principles to apply to a U-tube setting.

Table 3: Order of major steps in solution to the U-tube problem by (a) ABLE, (b) and (c) More able human solvers, (d) and (e) less able human solvers, (f) Text example.

(a) ABLE

$p_C(original) = p_C(added)$
$p_A(original) = p_A(added)$
$\Delta p_{original} = \Delta p_{added}$
$l = l_{up} = l_{down}$
$h_{original} = 2l$
$\Delta p_{original} = \rho_m g2l$
$\Delta p_{added} = \rho_w g(13.6\ cm)$
$\rho_m g2l = \rho_w g(13.6\ cm)$
$l = 0.486\ cm$

(b) S1

$p_C(original) = p_C(added)$

$l_{up} = l_{down}$
$\Delta p_{original} = \rho_m g2l$
$\Delta p_{added} = \rho_w (13.6\ cm)$
$\rho_m g2l = \rho_w g(13.6\ cm)$
$l = 0.486\ cm$

(c) S2

$l_{up} = l_{down}$

$\Delta p_{original} = \Delta p_{added}$
$p_A(original) = p_A(added)$

$h_{original} = 2l$
$\Delta p_{original} = \rho_m g2l$
$\Delta p_{added} = \rho_w g(13.6\ cm)$
$\rho_m g2l = \rho_w g(13.6\ cm)$
$l = 0.486\ cm$

(d) S3

$p = p_0 + \rho gh$
(decide U-tube example is more relevant)
$\rho_w / \rho_m = 2l/(2l + d)$
$l = 13.6\ cm$
[algebra and arithmetic]
$d = 353.6\ cm$
(too large try another approach)
$d = 13.6\ cm$
[algebra and arithmetic]
$l = 0.523\ cm$

(e) S4

$p = p_0 + \rho gh$
h = height of mercury column
$p = p_0$ = atmospheric pressure
(no place for desired quantity try another approach)
$p = p_0 + \rho g(y_2 - y_1)$
$y_2 = 13.6\ cm$
y_1 = height of mercury column
ρ = density of mercury
$p = p_0$ = atmospheric pressure
$y_2 = 1.35 \times 10^{-1}$ meter

(f) Text Example

$p_C(added) = p_C(original)$
$\Delta p_{added} = \Delta p_{original}$
$p_A(added) = p_A(original)$
$\Delta p_{original} = \rho_w g2l$
$h_{original} = 2l$
$l = l_{up} = l_{down}$
$\Delta p_{added} = \rho g(d + 2l)$
$\rho_w g2l = \rho g(d + 2l)$
$\rho / \rho_w = 2l/(2l + d)$

The order in which principles are applied is similar to that generated by the ABLE system, and differs slightly from the order presented in the original example in that information is generated in a forward working manner so that information is always available at the time it is needed. These data are presented in Table 3. Part (f) of Table 3 shows the nine inferences, labeled 1-9 in Table 1, used by the textbook in solving the analogous problem stated in Table 1. The remaining solutions are for the problem solved by the human subjects.

The two subjects considered here both solved this problem correctly. All of the subjects we consider to be more able either solved this problem correctly or made the simple error of solving for the height of the original fluid (mercury) above the point C (Figure 1a), rather than solving for half that distance, the distance the mercury rose.

2.2. Less able Learners

Reading

Unlike ABLE and the more able solvers, the less able solvers skimmed through the text rapidly. The two considered here

complained that reading aloud interfered with understanding and were permitted to read the text silently, which they did rapidly (see Table 2). These subjects also both stopped reading and began solving the first problem as soon as they encountered material relevant to it.

Problem Solving

Unlike the more able subjects, the less able subjects do search through the text for appropriate relations. As Table 3 shows, S3 and S4 each tried two different principles. In all cases the selection was preceded by an episode of searching the text material.

As shown in Table 3 the order in which principles are applied by the less able subjects is very different from that produced by the more able subjects, by the ABLE system, and in the analogous text example. The procedure of these subjects seems to be the following: (1) search through the text for an equation that involves distances (presumably because a distance is the quantity to be found). This equation may be the equation resulting from the U-tube example (subject S3 in Table 3) but it may not be (subject S4). (2) Substitute values for quantities appearing in this equation, using as a criterion for substitution merely left over that the value substituted must be of the same type as the symbol for which it is substituted (i.e., a height for a height, a density for a density). Indeed, even this simple constraint is sometimes violated when subjects fail to distinguish between pressure p and density ρ. (3) If after this substitution all values in the quantity have been used, and there is in the equation a quantity of the appropriate type (here distance) left over, then solve the equation if possible. (4) If it is impossible to fit all the information in the problem into the equation, then abandon it and get a new one. (5) If there remain in the equation symbols not assigned values, then search either for an expression involving this symbol, or for some "standard" value for this symbol (e.g., atmospheric pressure).

This procedure is very different from that executed by ABLE. However, it is not hard to see how such performance might be produced by ABLE through appropriate deletion of strategic knowledge. First, one would have to use the initial ABLE system, before it had built any procedural knowledge about applying principles. This absence corresponds to the lack of processing of the text observed in these human solvers. Second, one would have to remove from ABLE its strategic knowledge that a principle can be applied only if all aspects of the setting of that principle are matched against the setting of the problem (production P4 in Figure 3). This production would be replaced with one that would allow use of a relation if all symbols in it could be matched by quantities of the same type in the problem by quantities of the same type (i.e., a length for a length). This uncritical matching perhaps is associated with the less able subjects' poor abilities for selecting useful principles. They have to match a lot of principles, and so may do it in a less costly way, even though this economy has devastating effects on their problem solutions. With these changes the ABLE system could produce any of the incorrect solutions we have observed in the less able subjects.

The result is a weak means-ends procedure of searching for relevant principles observed elsewhere in novice solvers (Simon and Simon, 1978, Larkin, McDermott, Simon, and Simon, 1980b). A first principle is proposed because it contains a quantity of the type to be solved for. Subsequent principles are proposed because they can be used to replace in the original equation quantities without known values. Substitution is based on the weak criterion that the two quantities must be of the same type (e.g., two lengths, two pressures).

Because of their uncritical matching of principles to the problem situation, the errors made by the less able subjects are varied and exotic compared to the simple "sensible" error characteristic of the more able subjects. The most common error is illustrated by subject S3 in Table 3. The equation from the U-tube example is used, the distance 13.6 cm is substituted for one of the distances in the equation, l and d, and the equation is solved for the remaining symbol for distance. However, as illustrated by subject S4, the errors can be far more exotic. However, all errors are produced by copying some equation from the problem, substituting for the symbols in that equation values that correspond in type, and then solving for a symbol for distance.

3. Conclusion

As we have noted elsewhere (Larkin, 1981a, Simon and Simon, 1978) individuals trained in physics seem to work with mental models that are different than those used by less trained individuals. In particular, skilled individuals re-represent the problems in terms of technical entities (e.g., pressure drops) that have no special meaning outside the discipline of physics. Here we suggest that the general learning mechanisms of proceduralization and composition provide some explanation of how this ability to re-represent problems might be acquired.

Our prototype ABLE system acquires a principle in declarative form, as a student might by reading a chapter. This initial encoding does not itself include any information about how or where to apply the principle. Thus initial applications of the principle are *interpretive*, achieved through general procedural knowledge about how to apply any principle or definition. Through such application ABLE builds new specific procedural knowledge associated with the principle. Initially fragments of procedural knowledge aid in the search process. They contain patterns of information that have been used with that principle in the past, thus short-circuiting ABLE's original general and weak method of selecting principles. Ultimately a principle can be completely proceduralized (for a set of analogous contexts) so that application is completely automatic. This final automatic knowledge may well be an ingredient of what one would want to call an expert's mental model in which technical entities (e.g., pressure drops) are seen as readily as visible entities like heights.

References

Clement, J. Students' preconceptions in introductory mechanics. *American Journal of Physics*, 1981, , in press.

Forgy, C.L. *Implementing a Fast Syntactic Pattern Matcher.* Technical Report, Computer Science Department, Carnegie-Mellon University, August 1979.

Forgy, C.E. *Preliminary OPS5 Manual.* Technical Report, Computer Science Department, Carnegie-Mellon University, 1980.

Green, B., McCloskey, M., Caramazza, A. Curvilinear Motion in the Absence of External Forces: Naive Beliefs About the Motion of Objects. *Science*, December 1980, *210(5)*, 1139-1141.

Halliday, D. and Resnick, R. *Fundamentals of Physics.* New York: John Wiley & Sons 1970.

Hayes, J.R., and Simon, H.A. Understanding written task instructions. In Gregg, L.W. (Ed.). *Knowledge and Cognition,* Hillsdale, NJ: Lawrence Erlbaum Assoc., 1974.

Larkin, J.H., McDermott, J., Simon, D.P., Simon, H.A. Models of competence in solving physics problems. *Cognitive Science*, 1980, *4*, 317-345.

Larkin, J.H., McDermott, J., Simon, D.P., and Simon, H.A. Expert and novice performance in solving physics problems. *Science*, June 1980, *208*, 1335-1342.

Larkin, J.H. *The Role of Problem Representation in Physics.* C.I.P. 429. Carnegie-Mellon University, 1981. Presented at a Conference on Mental Models, La Jolla, CA, October, 1980.

Larkin, J.H. Enriching formal knowledge: A model for learning to

solve problems in physics. In J R. Anderson (Ed), *Cognitive skills and their acquisition*, Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc., 1981.

Lewis, C.H. *Production system models of practice effects*. PhD thesis, University of Michigan, 1978.

McDermott, J. Some strengths of production system architectures. In *NATO A.S.I. proceedings: Structural/process theories of complex human behavior*, The Netherlands: A.W. Sijthoff, International Publishing Company, 1978.

Neves. D. and Anderson, J.R. Knowledge Compilation: Mechanisms for the Automatization of Cognitive Skills. In Anderson, J.R. (Ed.), *Cognitive Skills and Their Acquisition.*, Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1981.

Newell, A. and Simon, H. A. *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall, Inc. 1972.

Newell. A.. Shaw, J.C., & Simon, H.A. A variety of intelligent learning in a general problem solver. In Yovits, M.C. & Cameron, S (Eds.), *International Tracts in Computer Science and Technology and Their Application*, : Pergamon Press, 1960.

Newell, A. Production systems: Models of control structures. In Chase, W.G. (Ed.), *Visual information processing*, New York: Academic Press, 1973.

Simon, D.P. and Simon, H.A. Individual differences in solving physics problems. In Siegler, R. (Ed.), *Children's thinking: What develops?*, Hillsdale, NJ.: Lawrence Erlbaum Associates, 1978.

SUBMITTED PAPERS

Interestingness and Memory for Stories

John B. Black, Steven P. Shwartz
and Wendy G. Lehnert

Yale University

Abstract

Two experiments investigated the effects of the number of interesting events in a story on memory for that story. The results differed depending on whether or not the story followed a stereotypical script. For script-based stories, adding interesting events spaced throughout the stories decreased the stories' memorability. However, for non-script-based stories, adding a few interesting events aided memory for the stories, but adding too many decreased memory. These results were interpreted in terms of a limited resource model of story understanding that focusses processing on interesting events.

Introduction

One of the major problem areas for theories of language understanding is how to determine which inferences are important to make while understanding a story and which inferences are not important to make. When inferencing is left uncontrolled, a combinatorial explosion of inferences is often the result.

Schank and Abelson (1977) proposed a group of higher level knowledge structures (scripts, plans, goals and themes) that constrain inferencing to only the most relevant paths, but Schank (1979) concluded that even more constraint was needed. In particular, he proposed that interestingness be used to guide inferencing: i.e., that the inferencing process should concentrate on the most interesting events in a narrative. Schank argued that an interestingness value could be assigned to events in narratives using the criteria that some events are inherently interesting (e.g., sex, death and violence are always interesting) and other events are interesting in certain contexts (e.g., taking off one's clothes in public is interesting but not in private). These interestingness values can then be used to focus the inferencing processes on only the most interesting events in a narrative.

Lehnert (1979) extended this analysis by adding the notion of resource limitations and making predictions for the recall of stories. One of the predictions was that adding interesting events to a story will improve its memorability because they provide markers around which events in the story can be organized. However, because processing resources are limited, adding interesting events to a story will only aid memory up to a certain point. If a story contains too many interesting events then there will be too many high priority events competing for the limited processing resources, so this "overload" condition should also be detrimental to recall of the story. We conducted two experiments to test these predictions.

Number of Interesting Events Experiment

In this experiment we explicitly tested the prediction that as the number of interesting events in a story increased, the memory for the story would first be aided and then be hurt. In particular, we designed four stories of 20 events each such that up to nine unrelated but inherently interesting events could be substituted for uninteresting events with minimal change in the sentences. For example, one story about a visit to a library contained the following two uninteresting events

> On her way Mary passed a friend
> and remembered she was in her class.

The first sentence here can be made interesting by merely substituting "prostitute" for "friend."

Two of these four stories were based on stereotypical scripts (visiting the library and eating at a restaurant) while the other two were not (walking in the park and visiting a private club). We wrote four versions of each of these stories. These versions differed only in the number of interesting events. One version had 0, another had 3, a third had 6 and the final one 9. These interesting events were distributed evenly throughout each story version. Our main interest was in how this number of interesting events would affect the subjects' recall of the stories. We gave the four stories (one in each version) to 32 subjects to read (8 subjects saw a given version of a given story), then after a 15 minute intervening task the subjects were given the story titles and asked to recall as much as they could from the stories. These recall protocols were then scored for the gist of the events in the stories.

Table 1 gives the percentage of the story statements recalled for each version of the stories.

Table 1

Percent of Story Statements Recalled in First Experiment

Type of Story	Number of Interesting Events			
	0	3	6	9
Script	65	51	52	56
Non-script	53	59	49	42

The results were quite different for the different kinds of story. In particular, the non-script stories followed the predicted pattern: i.e., adding 3 interesting events improved recall but adding more hurt recall. However, recall of the script stories was best with 0 interesting events, second best with 9, and worse with the intermediate values of 3 and 6.

Our interpretation of these results is that in the script-based stories there is already a story organization (namely, the script), so adding interesting events is harmful to recall because the existing organization is disrupted. In the 0 interesting events condition, subjects can reconstruct a script-based story during recall by merely remembering that it was a typical implementation of the script (e.g., a typical library story). However, adding interesting events makes straight reconstruction no longer possible. The slight upturn at 9 interesting events might occur because by this time the subjects have completely abandoned trying to use the script as a reconstruction aid and merely focused on the interesting events as memory organizers.

With the non-script stories, on the other hand, the results were as predicted because there was no strong preexisting organization in the interesting events condition. Specifically, adding 3 interesting events aided recall, but the 6 and 9 events conditions led to worse recall because of an overabundance of interesting events.

Massed and Spaced Interesting Events Experiment

In this experiment we explicitly tested the prediction from Lehnert's limited resource model that it is the density of interesting events rather than the absolute number that determines recall. In the first experiment, the number of interesting events in the story was confounded with the density of interesting events. In this experiment we used the same basic stories but had only the 0 and 3 interesting events conditions. Now, however, we varied density independently of number by having the interesting events either be located adjacent to one another (the massed condition) or be evenly spaced throughout the story as before (the spaced condition). The procedure was the same as in the first experiment with 24 subjects this time (but still 8 per group).

Table 2 gives the percentage of the story statements recalled in the versions of the stories.

Table 2

Percent of Story Statements Recalled in Second Experiment

Type of Story	Number of Interesting Events		
	0	3 Spaced	3 Massed
Script	55	58	64
Non-script	50	55	46

As with the first experiment, the results were quite different for the two types of stories. With the script stories, there was essentially no difference between the 0- and 3-spaced conditions, but the 3-massed condition led to better recall of the story. With the non-script stories, on the other hand, the results confirmed the expectation that the spaced interesting events would improve recall, while the massed ones would decrease recall.

Our interpretation of these results is that with the script-based stories, having the disruptive interesting events massed allows the reader to store them in memory as separate units which is harder to do if the interesting events are spaced throughout the story. Thus the massed condition facilitates a memory representation like the one proposed by Graesser, Gordon, and Sawyer (1979) -- namely, remembering the story as a script plus a list of deviations.

With the non-script stories, on the other hand, such a representation is not possible because there are no organizing scripts. In these stories, the interesting events provide the organization for the story and the construction of this organization during reading is facilitated by having the interesting events spaced throughout the story. If the interesting events are massed, the limited resources available for processing are "locally overloaded" so the memory for a story with massed interesting events is even worse than one with no interesting events.

Conclusions

The non-script stories conformed to our expectations that adding a few interesting events spaced throughout a story would increase the story's memorability, but that adding too many interesting events or having them massed together would decrease the story's memorability. The results were different, however, if the stories were highly stereotyped, script-based stories. In particular, adding interesting events to script-based stories disrupts the existing script organization and hence led to less memorable stories unless these disruptive events were massed together in the story. Thus a writer can "spice up" a text by adding inherently interesting statements, but this procedure will increase the reader's memory only if the text was poorly organized originally.

References

Graesser, A., Gordon, S.E. and Sawyer, J.D. Recognition memory for typical and atypical actions in scripted activities: Test of a script pointer + tag hypothesis. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 319-332.

Lehnert, W.G. Text processing effects and recall memory. Research Report no. 157, Department of Computer Science, Yale University, 1979.

Schank, R.C. Interestingness: Controlling inferences. Artificial Intelligence, 1979, 12, 273-297.

Schank, R.C. and Abelson, R.P. Scripts, plans, goals and understanding. Hillsdale, NJ: Erlbaum, 1977.

This research was supported by a grant from the Sloan foundation.

Use of Goal-Plan Knowledge In Understanding Stories

Edward E. Smith & Allan M. Collins
Bolt Beranek and Newman Inc.

There seems to be a growing consensus among researchers that understanding a story involves constructing a representation of the characters' goals and plans (e.g., Rumelhart, 1977; Schank & Abelson, 1977; Wilensky, 1978). There is, however, relatively little experimental evidence for people's on-line construction of such goal-plan representations, and the purpose of the present experiment was to provide such evidence.

In our paradigm, subjects are presented a five-line story, one line (or sentence) at a time. Subjects press a button as soon as they understand a sentence, and the sequence of reading times obtained with a particular story is assumed to offer a line-by-line description of the process whereby a representation is constructed for that story. The general idea behind our study was to systematically delete mention of certain goals and plans needed to understand the story, and to see if readers then took longer at lines where they would have to infer these missing goals and plans.

To get more specific, we have to deal with a story in detail. Consider then the following story about an operation.

1. While John was laid off, his wife became seriously ill.
2. John needed money to pay for an operation.

3a. He decided to borrow money from his Uncle Harry.

3b. He would give his Uncle Harry a quick call.

3c. He had to find out Uncle Harry's phone number.

4. John reached for the most recent suburban directory.
5. John was overjoyed Uncle Harry agreed to the loan.

By our analysis, this story can be represented by four levels of embedded goals and plans, where each level contains one goal and its associated plan. At the first or top level would be John's goal of his wife getting better (an inference from line 1) and his plan of getting her an operation (inference from line 2). A precondition for the "operation" plan, however, is that the person have money, and this precondition becomes the goal at the second level (it is explicitly stated in line 2), while the plan at the second level is to borrow money (see line 3a). The "borrow" plan also has a precondition, namely being in contact with the lender, and this becomes the goal at the third level (see line 3b); the plan at this level is to phone the lender (see line 3b). Finally, the "phone" plan has as a precondition that one know the telephone number, and this is the goal at the fourth level (see line 3c); the plan at this level is to use the directory (see line 4).

None of our subjects saw this full version of the operation story. Rather they saw a version with two lines deleted. As an example, some subjects read the story with lines 3a and 3b deleted. We expected these subjects to be relatively slow in reading line 3c because there is a sizeable gap between the goal mentioned in it and the goal and plan given by the directly preceding line 2. That is, by deleting lines 3a and 3b, we gapped a couple levels of goals and plans and our subjects must fill this gap when reading line 3c, which should take time. When these subjects get to line 4, however, they should read it relatively quickly; for now there is no gap in the underlying goal-plan representation that needs to be filled. In contrast to the case just described, other subjects read the operation study with lines 3b and 3c deleted. These subjects should read line 3a relatively quickly (there is no gap to fill in the goal-plan representation), but read line 4 relatively slowly (two levels of goals and plans have been gapped, and need to be filled at this point).

So, by varying which lines of the story are deleted, we can manipulate the size of the gap that a reader must fill when reading a particular story line. Since these gap-sizes are measured in units of underlying goals and plans, a finding that reading time increases with gap size would be evidence for the on-line construction of a goal-plan representation.

With stories like the operation one, we have obtained the predicted gap-size effect -- the time to read a line increases monotonically with the number of goals and plans that have to be filled in at that point -- though its magnitude tends to be greater on some lines than others. We were concerned, however, that the gap-size effect might only occur with stories that explicitly emphasize characters' intentions, i.e., readers might only construct goal-plan representations for stories that explicitly use intentional constructions like would give and decided to. For this reason (and others) we rewrote our stories so that sentences that were once intentional in tone now became actional. This required changing the order of the rewritten sentences, since the chronological order for actions is the reverse of that for intentions. In the actional version of the operation story, for example, John first reaches for the directory, then finds Uncle Harry's new number, then calls him, and then asks him for a loan -- the opposite order of lines 3a, 3b, 3c, and 4 in the original story.

Once having created these actional versions of the stories we then varied which lines were deleted, thereby varying the size of the goal-plan gap that the reader had to fill. We again found that the time to read a sentence increased with its associated gap size (though again the effect's magnitude depended on the exact line being read). These results suggest that people construct roughly the same goal-plan representation of a story regardless of whether the story emphasizes the characters' actions or intentions.

References

Rumelhart, D. E. Understanding and summarizing brief stories. In D. LaBerge & J. Samuels (Eds.), Basic processes in reading: Perception and comprehension. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Schank, R., & Abelson, R. Scripts, plans, goals, and understanding. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Wilensky, R. Why John married Mary: Understanding stories involving recurring goals. Cognitive Science, 1978, 2, 235-266.

INFERENCES IN STORY COMPREHENSION

RAYMONDE GUINDON
Department of Psychology
University of Colorado, Boulder

Abstract

Predictions were made about the types and the number of inferences to be found in the verbal protocols of subjects reading difficult to understand texts. The predictions were made on the basis of two alternative models of inference generation, the backward bottom-up inference and the forward bottom-up inference. It was also hypothesized that plan-goal and script inferences would be stored longer in STM than coreference and role identification inferences. This implies that plan-goal and script inferences are more likely to be reported than coreference and role identification inferences. The protocol analyses support the backward bottom-up inference model and support the assumed length of storage in STM of the different types of inferences.

1.0 Introduction

The study of the generation of inferences in text comprehension is important because it is believed to be one of the main mechanism by which a cohesive text representation is built. While inferences are not to be equated with expectations, it has not always been clear how inferences interact with top-down processing.

Some researchers have proposed a forward bottom-up inference model. The inferences generated at the input of a sentence are strictly determined by the local information in the sentence and they are unconstrained by the context (Rieger, 1975; Thorndyke, 1976). In Thorndyke's model (1978), diverse forward bottom-up inferences are generated from each new input sentence. Their number depends on the presentation rate of the text, the difficulty of the text, the purpose in reading, etc. Some of these inferences could then be compatible with an incoming sentence, facilitating its comprehension.

Other researchers have proposed backward, constrained models (Haviland and Clark, 1974; Kintsch and van Dijk, 1978; Wilensky, 1978). The inferences generated from the input sentence are constrained by the previous sentences. The inferences are produced specifically to establish a coherence relation between the semantic representation of the current sentence and the representation of the text. The type and number of inferences generated are constrained by the context.

Consider a simplified illustration of Wilensky's algorithm to explain events (1978). The inferences generated from an event are possible plans for that event. Those plans are checked to see whether one of these plans is a known plan, or is a plan for a know goal, or is a plan for a known theme, or is an instrumental plan for a know plan. Constraints are provided by already asserted or inferred themes, goals or plans, and by the fact that actions must be explained relative to them.

Experimental evidence seem to be somewhat more in favor of a backward, constrained generation of inferences. The results of Thorndyke's experiment (1976) were proposed as a support for a forward bottom-up inference model. However, this experiment might suffer an identifiability problem. Using previously experimentally derived inferences, a recognition test with those inferences and sentences of the texts is given. It is found that the false alarm rate for compatible inferences is greater than for incompatible inferences. However, such predictions can be derived as well from a backward bottom-up inference model than from a forward bottom-up inference model. A backward inference model will predict that the compatible inferences, that is, those which establish coherence, become part of the text representation and are likely to be falsely recognized as text sentences (see Keenan,McKoon and Kintsch, in Kintsch, 1974). Incompatible inferences, which do

not establish coherence, do not become part of the text representation and are not likely to be falsely recognized.

In an experiment by Miller and Kintsch (1980), subjects have to produce continuations for a sentence presented by itself, in a specific context establishing clearly a main topic, or in a non-secific context suggesting many topics. It is predicted and found that the subjects in the no or non-specific context will rely on the local constraints provided by the target sentence to produce these continuations. On the other hand, it is predicted and found that subjects in the specific context condition will produce continuations related to the main topic. Any model of inference generation, backward or forward, will predict the continuations to be determined by the local constraints in the no or non-specific contexts. However, a forward inference model cannot account for the fact that the continuations produced in the specific context condition are all related to the main topic (constrained), while a backward inference model predicts just that.

2.0 Method and Predictions

The study to be presented makes predictions over the types and the number of inferences likely to be found in the verbal protocols of subjects reading difficult-to-understand texts. These predictions are based on the two alternative models of inference generation, the forward and the backward models, they are based on assumed or known properties of the memory storage of inferences, and they are based on the theory of verbal reports (Ericsson and Simon, 1980).

Nine subjects were trained to make thinking-aloud and retrospective reports on their inferences on twelve practice texts. It was emphasized that they should not try to report all the inferences that they could eventually make, but only report the inferences they were making automatically, without effort. However, they were also instructed that some texts might be harder than others and that it would be normal to find the production of inferences more difficult in those cases, but still to report them.

The texts used were "Paul's Outing" (Collins, Brown and Larkin, 1977), "The Crowd", an adaptation of a text used by Bransford and Johnson (1973), and "Noon, Downtown", a short humourous narrative. In all the texts, the last sentence indicated clearly the topic of the text. The text "The Crowd" is more descriptive than narrative.

For an inference to be reported, it must be stored in STM and during a minimal amount of time (Ericsson and Simon, 1980). Inferences likely to be reported are inferences of plans and goals, and inferences of scripts (Schank and Abelson, 1977). In the protocols, explicit inferences of plans and goals would be like: "The driver went underneath the truck to repair it.". Explicit inferences of scripts would be like: "This is going to be about traffic." or "He seems to go to a movie.".

Inferences of plans and goals are likely to be reported because they are the major coherence relations in the text representation of narratives. They need to be stored long enough in STM (short term memory) to build the cohesive text representation and they become a permanent part of the text representation. Inferences of scripts are likely to be reported because the scripts are used to interpret a certain number of sentences and are the basis of the text representation.

Inferences not likely to be reported are coreference relations and role identifications. Coreference relations are the determination of the referent of a pronoun or a definite description. They are often made automatically, on the basis of the structure of the text, without requiring STM to store intermediate computations, and therefore, are not likely to be reported. Role identifications are the recognition that a story character or object correspond to a script character or object. The recognition is made on the basis of functional or categorical information provided by the text. Presumably, this is accomplised through direct pattern-matching and therefore does not use STM to store intermediate results and cannot be reported. However, if the inferred coreference or role identification reveals itself to be inadequate, it is likely that the reader will report the inadequacy and will report the new assignment.

For each text, a set of inferences were a priori derived and predicted to appear explicitly, either frequently or infrequently, in the protocols.

It is predicted that the most frequent explicit inferences will be those of plans and goals, and scripts. A related prediction is that the most frequently implicit inferences will be those of coreferences and role identifications. By implicit, it is meant that the protocol in icates rather certainly that the inference was made but only indirectly. For example, suppose two of the sentences were: "Paul plunked down $5 at the window....but he refused to take it.". Suppose also that the protocol contains something like: "Well, I couldn't understand why John wouldn't take the change.". It is clear that the reader made the coreference relation between "he" and "John".

According to the backward bottom-up inference models, plan-goal inferences should be reported only after the event they help to explain is read, and in most cases, only one coherence relation should be reported for that event. According to the forward bottom-up inference models (Thorndyke, 1976), numerous inferences could be reported when a sentence is read, potentially explaining an event in a future sentence.

The protocols were analysed in terms of the types and frequency of the explicitly and implicitly reported inferences. A simple consistency checking was performed on the classification of the inferences using a third of the protocols (9 out of 27). The consistency estimate was about .95.

In every case of a reported coherence relation, it has been reported after the event it explained, and in only one case has there een ore than one coherence relation proposed (there were two). This is consistent with a backward bottom-up inference model.

Table 1 presents the types, predicted frequencies and observed frequencies in the protocols of the a priori determined inferences.

The overall frequency of the predicted frequent inferences is .65, and of the unfrequent ones is .11.

As was predicted, plan-goal and script inferences are much more ferquently reported than coreference or role identification inferences.

Table 2 presents the frequencies of inferences explicitly generated, implicitly generated and of inferences for which there is no indication in the protocols that they have been made. The data are summed over the two methods because while there is a slight tendency for more inferences to be reported in the thinking-aloud reports, this is not true for all stories.

Script and plan-goal inferences are more often explicit than implicit or not mentionned, while coreference and role identification inferences are most often implicit.

The location and the number of inferred coherence relations support a backward bottom-up inference model. The greater frequency of reported plan-goal and script inferences is consistent with their assumed length of storage in STM and LTM. The relative unfrequency of role identification and coreference inferences supports their assumed automaticity and lack of use of STM to store intermadiate results.

The somewhat less good fit of the data to the predictions in "The Crowd" might be due to the fact that "The Crowd' is more descriptive than narrative anc that the models for inference generatior were devised for narratives.

Table 1
Observed frequencies of each instance of the predicted inferences summed over the two methods

NOON, DOWNTOWN (each on a total of 5)

Predicted frequent:

p-g	cause	p-s	p-s	p-g	p-g	p-g	
5	3	4	1	4	2	3	M= .6

Predicted unfrequent:
None of the 6 instances of coreferences has been observed in any protocol M= .0

PAUL'S OUTING (over a total of 6)

Predicted frequent:

p-g	p-g	p-g	p-g	
5	5	6	5	M= .9

Predicted unfrequent:

cor	cor	role	role	role	role	cause	
1	2	0	1	0	1	0	
role	cause						
0	2						M= .1

THE CROWD (over a total of 6)

Predicted frequent:

scr	scr	p-g	scr	scr	cause	scr	
1	3	4	2	4	1	3	M= .5

Predicted unfrequent:

role	role	cor	role	cor			
1	1	0	1	0			
impl	impl						
2	1						M= .2

p-g = plan-goal role = role-identification
scr = script impl = implicature
M = mean

Table 2
Types and frequencies of inferences as a function of stories and methods.

NOON, DOWNTOWN

EXPLICIT	IMPLICIT	NOT FOUND
21 p-g	6(2) p-g	3(5) p-g
0 cor	12 cor	

PAUL'S OUTING

EXPLICIT	IMPLICIT	NOT FOUND
20 p-g	3(2) p-g	1(0) p-g
3 cor	9 cor	
2 role	12 role	
1 cause		

THE CROWD

EXPLICIT	IMPLICIT	NOT FOUND
16(15) p-g	7 role	11(12) p-g
5(4) role	7 cor	
3 impl	1(0) p-g	
1 cause		
1(0) cor		

Acknowledgment

I would like to thank Dr. Anders Ericsson for his useful advice and encouragement.

References

Bransford, J. D., Johnson, M. K. Contextual Prerequisites for Understanding: Some Investigations of Comprehension and Recall. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 717-726.

Collins, A. M., Brown, J. S., Larkin, K. M. Inference in Text Understanding. Technical Report No. 40. Bolt, Beraneck and Newman, 1977.

Ericsson, K. A., Simon, H. A. Verbal Reports as Data. Psychological Review, 1980, 87, 215-251.

Kintsch, W. The Representation of Meaning in Memory. Hillsdale, N.J.: Erlbaum, 1974.

Kintsch, W., van Dijk, T. A. Toward a Model of Text Comprehension and Production. Psychological Review, 1978, 85, 363-394.

Miller, J. R., Kintsch, W. Knowledge-based Aspects of Prose Comprehension and Readability. Presented at the annual meeting of the American Educational Research Association, Boston, April 1980.

Rieger, C. Conceptual Memory and Inference. In R. C. Schank (Ed.), Conceptual Information Processing. Amsterdam: North-Holland, 1975.

Schank, R. C., Abelson, R. P. Scripts, Plans, Goals, and Understanding. Lawrence Erlbaum, 1977.

Thorndyke, P. W. The Role of Inferences in Discourse Comprehension. Journal of Verbal Learning and Verbal Behavior, 1976, 15, 437-446.

Wilensky, R. Understanding Goal-based Stories. Research Report No. 140, Department of Computer Science, Yale University, 1978.

Self-Embedding is Not A Linguistic Issue*

John M Carroll
MIT Linguistics and Philosophy Dept.
and IBM Watson Research Center

Anyone familiar with language research over the last 25 years or so will be no stranger to examples like these:

The woman died.
The woman the man met died.
The woman the man the girl loved met died.

Such examples have an illustrious past, both in the formal study of language and in the psychology of language. The self-embedding (SE) property they display singularly places natural language syntax beyond the generative capacity of finite devices (Chomsky, 1959).** The interaction of SE with language comprehension entrains a remarkable phenomenon: one (or zero) levels of SE cause no noticable increment in comprehension difficulty, but two (or more) levels are typically associated with substantial impairment of comprehension (e g., Miller and Isard, 1964).

Explaining SE. This remarkable fact has maintained the study of SE as a preeminent topic in psycholinguistics. A variety of accounts have been produced The earliest accounts tended to emphasize the extent to which SE overtaxed memory resources for speakers and hearers (Miller and Chomsky, 1963: 470ff.; Yngve, 1960). Later accounts tended to focus on the kinds of operations that were invoked in the course of processing SE structures (Miller and Isard, 1964; Bever, 1970). Significantly, though, all of these accounts regarded SE as a general structural property -- whose expression in language was of special interest -- not as a property unique to language ex hypothesi. Miller and Chomsky (1963: 484) write "Self-embedding of such great theoretical significance ... that we should certainly look for occurrences of it in non-linguistic contexts." Bever (1970) cited as one of the strengths of his account of SE processing problems that analogous principles seemed to explain phenomena in visual perception.

Clearly though, there is another theoretical option. One could hypothesize that SE phenomena are special to language. On such a view, the natural explanatory mechanism for the interaction of SE with sentence comprehension would be language-specific -- not some general property of memory or perceptual process as suggested by Bever, Miller and Chomsky, etc. This view is (often implicitly) adopted by many natural language parsing theorists. A recent example is Fodor and Frazier (1980) who suggest that the interaction of SE with sentence comprehension be theoretically reconstructed as a parsing principle attaching an incoming word or phrase into its surface structure using the smallest possible number of new nonterminal nodes (Fodor and Frazier, 1980: 426-434). (See Carroll, 1981; Ford, Bresnan, and Kaplan, 1981, and Wanner, 1980; for further discussion of this model.)

The question of whether generalizations about SE phenomena are linguistic or still more general is a question of fact. But it is worth emphasizing that assumptions one way or the other lead immediately to empirical consequence. For example, the assumption that SE is strictly linguistic actually does some work in the analysis of Fodor and Frazier. The single strong distinction Fodor and Frazier (1980) are able to draw between their model and that of Wanner (1980) turns precisely on their language-specific analysis of SE phenomena.

While the resolution of this issue may ultimately be decided on theoretical grounds, I will focus chiefly on the presentation of data whose prima facie analysis entails the view that the correct level of generalization for addressing SE phenomena is something like "complex sequences" I will discuss examples from film, dance, music, and social interaction Then, I will turn back to the theoretical level and outline a proposal, following Bever's (1970) Double Function Hypothesis.

Film. An obvious candidate for SE analysis in film is the "flashback" scene: an entire scene is embedded into another scene. In the cinema of D.W. Griffith, for example, a scene (S) can consist of a long-shot (L) followed by a series of detail-shots (D) A detail-shot may itself include a scene. This has the effect of embedding a scene within a scene In Carroll (1980: 61-63) I have discussed the following grammar-fragment as the basis of a formal analysis of this sort of composition:

S ==> L + D*
D ==> D' + S + D'

The right-hand side of the second rule introduces the "prime" as a notation to indicate that a return is structurally implied. This amounts to a special sort of indexing

Several examples of this structure occur in Griffith's Broken Blossoms. Richard Barthlemess, as the Chinaman, casts a misty-eyed look within a detail-shot, and immediately there is a cut away to a flashback of his youth. It would not have been possible for Griffith to cut away from Barthlemess, show the flashback, and then proceed with the story-line without first returning to Barthlemess. The return is obligatory; it is part of the structure of (Griffith's) cinema. As subsequent examples will show this is the crucial property of sequences that affords SE. If a sequential domain has a structure strong enough to afford obligatory returns, it can have SE. (This, of course, is precisely analogous to points made by Chomsky, 1959, with respect to language.)

Dance. A well-known example of dance within dance is in the Nutcracker Suite where several episodes of puppet-dance occur as part of the ballet itself. More common, but also more subtle, are examples in which a dance phrase (Lasher, 1981) is interrupted, whereupon a complete and distinct other phrase is danced, then finally the original phrase runs to completion.

One class of such examples occurs when a female dancer is lifted by a male dancer in classical ballet: the female assumes a rigid posture while the male carries her and continues moving; when she is returned to the floor, the female continues with her own dance. Another class of examples are cases in which a group of dancers assume rigid postures while a subset, usually a couple or a single dancer, perform a dance. At the completion of the embedded dance, the group resumes. These latter types of examples have analogs in music where they are easier to cite because of the available notation.

Music. Music has an extremely articulated structure and affords several classes of SE structures. One of these is the cadenza: a strong cadence is interrupted by a solo and then completed. For example, the cadence I-6/4 -- V7 -- I, closing a section of a violin concerto, can be interrupted by a violin solo: I-6/4 -- V7 -- Solo -- I

When the interruption amounts to a delay in the matrix structure, it is often called "parenthesis" (Meyer, 1973). Meyer discusses an example from Haydn's String Quartet in Eb Major, Opus 50 No. 3: a melodic patterning of thirds Eb-G, F-Ab, implying G-Bb, and reinforced by harmonic and rhythmic patterns, is interrupted by a four measure repeating a-b-a-b structure: ". . the real melody is characterized by goal-directed motion; but the parenthesis is static."(p. 241)

Social interaction. For some years now, Sacks (1973) and associates have been developing a theory of the structure of conversational interactions based on the "adjacency pair" unit. The paradigm example is question/answer:

A: Do you know what an adjacency pair is?
B: No.

The implication of the "second pair part" entailed by the occurrence of the "first pair part" is sufficient structure to support SE Indeed, Goffman (1979: 258-9) noticed such examples.

A1: Can I borrow your hose?
B2: Do you need it this very moment?
A2: No.
B1: Yes.

Instead of answering A's initial question (and completeing an adjacency pair structure), B initiates a second question-answer adjacency pair. When this second pair is complete, there is a return to the first structure. Goffman also noted an example of multiple SE (B is a trainman in a station):

A1: Have you got the time?
B2 Standard or Daylight Saving?
A3: What time are you running on?
B3: Standard.
A2: Standard then.
B1: It's five o'clock

Jefferson (n.d.) has shown however that such cases of multiple SE are extremely rare, although noniterative SE is quite common.

Indeed, it is striking that, as in the case of language, multiple SE structures simply do not obtain One certainly has flashbacks, but never flashbacks within flashbacks When such structures are employed (as by Alain Resnais in Je T'aime, Je T'aime), they are employed precisely in order to confuse the viewer with respect to temporal sequence Analogously, one cannot imagine straightforward contexts in which a dance phrase within a dance phrase within a dance phrase could appear in a ballet One cannot embed a theme within a parenthesis within a theme One cannot embed a play within a play within a play Etc. In the final section of this paper I want to return to the theoretical analysis of SE phenomena and to suggest a general account of this intermodal restriction.

Double function. The double function principle can be put quite simply as: "The same stimulus cannot be perceived in two incompatible ways at the same time." The prima facie force of the principle in visual perception and language comprehension (Bever, 1970) and cinema perception (Carroll, 1980: 190-193), exclusive of SE phenomena, has been reviewed elsewhere. As applied to SE, the line one would want to run is that the categories "embedder" and "embeddee" are perceptually significant and that accordingly an object of perception cannot simultaneously be both embedder and embeddee. From this the usual facts regarding SE follow.

This remains an empirical hypothesis, of course, but the plausibility of the premises for the analysis is considerable. In contrast, appeals to memory (Miller and Chomsky, 1963; Yngve, 1960) are less compelling in that people typically can memorize multiple SE sequences -- they just can't understand them Appeals to processing constraint are either too vague to assess (Miller and Isard, 1964) or tantamount to the proposal under discussion (Bever, 1970). Finally, the cross-modal evidence of SE and the iterative SE constraint strongly indicate that language-specific analyses are missing the true generalization.

Notes

*This work was facilitated by a fellowship from the National Science Foundation and a sabbatical leave from the IBM Corporation. I am grateful to Tom Bever, Jeff Coulter, Len Meyer, and Margot Lasher for discussion and criticism.

**I will presuppose, but not review, the standard distinction between "nesting" and "self-embedding" (Chomsky and Miller, 1963).

References

Bever, T G. The cognitive basis for linguistic structures. in J.R. Hayes (Ed.) Cognition and language learning. New York: Wiley, 1970.

Carroll, J.M. Toward a structural psychology of cinema. The Hague: Mouton, 1980.

Carroll, J.M. On fallen horses racing past barns. In Papers from the parasession on language and behavior. Chicago: Chicago Linguistic Society, 1981.

Chomsky, N. On certain formal properties of grammars. Information and Control, 1959, 2, 137-167.

Chomsky, N. and Miller, G.A. Introduction to the formal analysis of natural languages. In Luce et al. (1963).

Fodor, J D. and Frazier, L. Is the human sentence parser an ATN? Cognition, 1980, 8, 417-459.

Ford, M., Bresnan, J., and Kaplan, R. A competence-based theory of syntactic closure. In J Bresnan (Ed.) The mental representation of grammatical relations. Cambridge: MIT Press, 1981.

Goffman, E. Replies and responses. Language in Society, 1979, 5, 257-313.

Jefferson, G. Unpublished research, n.d.

Lasher, M.D. The cognitive representation of an event involving human motion. Cognitive Psychology, 1981.

Luce, R.D., Bush, R.R., and Galanter, E. (Eds.) Handbook of mathematical psychology. New York: Wiley, 1963.

Meyer, L.B. Explaining music. Berkeley: University of California Press, 1973.

Miller, G.A. and Chomsky, N. Finitary models of language users. In Luce et al. (1963).

Miller, G.A. and Isard, S. Free recall of self-embedded English sentences. Information and Control, 1964, 7, 292-303.

Sacks, H. Lecture notes. Summer Institute of Linguistics, Ann Arbor, Michigan, 1973.

Wanner, E. The ATN and the sausage machine: Which one is balony? Cognition, 1980, 8, 209-225.

Yngve, V.H. A model and an hypothesis for language structure. Proceedings of the American Philosophical Society, 1960, 104, 444-466.

Center-Embedding Revisited

Michael B. Kac
Dept. of Linguistics
University of Minnesota
Minneapolis, MN 55455

The severe comprehension difficulty associated with certain center-embedding constructions is perhaps the best known of psychosyntactic phenomena. Most attempts at explanation have been variations on a single theme--that the c.e. configuration leads to an overload of short-term memory during processing. That this is not the whole story can be seen from considering the fact, rarely noted, that c.e. constructions exist which are understood quite easily, e.g.

(1) If either the Pope is Catholic or pigs have wings then Napoleon loves Josephine.

To this observation it might be replied that it is not c.e. per se that causes difficulty, but rather MULTIPLE c.e.; thus, (1) would not be expected to pose problems since it is embedded only to a depth of 1. But the same is true of

(2) If if the Pope is Catholic then pigs have wings then Napoleon loves Josephine.

which is, at best, at the outer reaches of comprehensibility.

The facts regarding (1-2) can be accounted for by a few simple assumptions. The first is that constituent recognition is carried out in strict left-right fashion; assume further that in attempting to analyze (1-2), the parser has at some point built a structure of the form

(3) If/either S_1 then/or S_2 then S_3.

and that at this point the following procedures are invoked:

(4) a. When an if is encountered, open an S at that point; the locate the first then to the right of this if and close at the end of the S immediately following.

b. When an either is encountered, open an S at that point; then locate the first or to the right of this either and close at the end of the S immediately following.

Applied to (1), (4) will operate in straightforward fashion; it will close the S beginning with if directly after S_3, and the S beginning with either directly after S_2. Applied to (2), however, it will misparse, closing the S beginning with the first if prematurely after S_2--a garden path effect of a familiar kind. A parallel account can be given of the difference between

(5) a. That for Harry to like Maxine would annoy Fred bothers me.

b. That that Harry likes Maxine annoys Fred bothers me.

where both involve nesting of complementizer-verb dependencies, but where (5b) is considerably more difficult to process than (5a). Assume that complementizers are associated with the verbs that they mark as non-main by the following procedure:

(6) a. Link each for and to to the first infinitive to its right.

b. Link each that to the first finite verb to its right.

In (5a), that is linked to would while the for and to are both linked to like, as desired; in (5b), however, there will be a garden path since the first that will be erroneously linked to likes rather than annoys. Thus (5a-b) and (1-2) are treated analogously in that that errors of prematurity are committed in the cases that pose comprehension difficulty but not in the cases that don't.

The general idea embodied in the foregoing can be further extended. Consider Object relative constructions like

(7) a. The rat that the cat chased squeaked.

b. The rat that the cat that the dog bit chased squeaked.

Examples of this type are perhaps most familiar from discussions of c.e. That (7b) should be more difficult to process than (7a) falls directly out of general design features of a syntactic parser currently under investigation called MULTIGAP ('multiple pass group-analyzing parser') and is, moreover, attributed to a garden path effect much like the ones hypothesized in the earlier cases discussed. In MULTIGAP, although simple NP's (i.e. NP's without clausal modifiers) are identified early in parsing, recognition of complex NP's is forestalled until after an exploratory phase during which the parser builds a structural representation called a PREANALYSIS, in which the sentence being parsed is parcelled up into a sequence of units called BOUNDED GROUPS (b-groups). A key notion here is that of the TRANSITION from one predicate to the next (or from the last predicate to #), i.e. the material that intervenes between the former and the latter. If conditions are satisfied for construing the latter as subordinate to the former (e.g. if there is an overt subordinator in the transition) then the transition is of one type, labelled B'; otherwise, it is of a different type, labelled B. Cross-cutting this dichotomy is a distinction between STRONG and WEAK transitions. If the transition from a predicate of n places to the next predicate, or to #, contains at least n-1 NP's, it is strong, weak otherwise. In (7b), the transitions from bit to chased and from chased to squeaked are both weak, while that from squeaked to # is strong. Abbreviatorily, the four transition types are labelled s (strong B), w (weak B), s' (strong B'), and w' (weak B'). The parser not only types transitions according to this scheme, but also inserts into each a special boundary marker,][, to delineate b-groups. If a transition is of any type other than s,][is positioned directly after the initial predicate in the sequence under consideration, otherwise directly after the last NP of the transition. (Recall that simple NP recognition has already taken place.) The preanalysis of (7b) is thus

(8) The rat that the cat that the dog bit][chased][squeaked][#
$\underline{w}_1$ $\underline{w}_2$ $\underline{s}$

Once the preanalysis is set, the parser looks for possible opening points of complex NP's--sequences of the form NP-SUB-V and NP-SUB-NP. (The full MULTIGAP design is capable of dealing with cases where no overt subordinator occurs, but discussion will be limited here to cases where there is such a subordinator.) The opening is labelled as belonging to type 1 (Subject relative) or type 2 (other) depending on which type of opening sequence is found. Accordingly, the parser will find two type 2 openings in (8): the rat that the cat and the cat that the dog. Closure is effected by a procedure which, if the opening is type 1, closes at the first s to the right, but at the first w to the right of a type 2 opening. Thus in (7a), the complex NP will be correctly closed after chased, but in (7b), there will be premature closure of the larger NP at $\underline{w}_1$ rather than at $\underline{w}_2$, an error of prematurity exactly analogous to those discussed earlier. Note, moreover, that c.e. of a type 1 relative in a type 2 construction poses no problems:

(9) The mouse that the cat who chased the rat saw squeaked.

While (9) is not absolutely straightforward, it can be understood with a little effort, which is not the case with (7b) even though (9) is of the same depth of embedding. The difference is accounted for in this treatment by the fact that the larger NP is closed off directly after saw (i.e. after the one w) while the smaller one closes directly after rat--at the first s. Because there is only one w, no error of prematurity occurs.

Complex NP's actually come in two varieties, which we might call first and second degree; in a first degree NP. there is only one predicate, while in a second degree NP the main predicate of the construction has a complement. The two types of NP are somewhat different in their behavior in that while type 2 NP's of the first degree always end in w-transitions, a type 2 NP of the second degree may end in either a w or an s; thus compare

(10) a. The boy that Harry believes][likes Maxine][saw Sue][# w'
$\underline{s}_1$ $\underline{s}_2$

b. The boy that Harry believes][Maxine likes][saw Sue][s'
w s

In (10a), the Subject NP ends at $\underline{s}_1$ while in (10b) it ends at w, corresponding to the fact that in the former case the head NP is the Subject of the complement clause while in the latter it is the Object of that clause. The parser copes with this fact by being equipped with two closure mechanisms, one for first degree complex NP's, and another for second degree NP's. A first degree NP can be identified by checking to see that no B' intervenes between the opening and the first B to the right thereof; if such a B' is found, closure is forestalled. Once all first degree NP's have been closed, the parser closes second degree NP's by simply looking for the first available B to the right of any opening for which no corresponding closure has been made. (Any B that has already been swallowed up into a previously recognized NP is no longer available for consideration.) It is a consequence of this feature of the parser that a second degree complex NP with a c.e. first degree NP should be more easily processed than one with a second degree NP, an expectation that is evidently borne out: compare

(11) The boy who believes that the girl who kissed Harry likes Maxine saw Sue.

which, like (9), requires some effort, but is within bounds; however,

(12) The boy who believes that the girl who thinks that Harry likes Maxine saw Sue kissed Samantha.

evidently is not. In the case of (11), the inner NP is parsed first, then the outer one, and there is no erroneous closure. In the case of (12), by contrast, there are no first degree NP's, and the NP beginning with the boy is closed prematurely, after Maxine.

Honesty compels mention of a perplexing case that this approach cannot handle. If, in a c.e. type 2 relative of the first degree, one of the NP's is replaced by a pronoun, comprehensibility seems to increase:

(13) The rat that the cat I saw chased squeaked.

I have no explanation to offer for this fact.

A general assumption underlies the treatment advanced here, as follows: in an optimal parser, one seeks the simplest procedures that will apply in error-free fashion to the simplest cases--cases in which a construction of type x contains no other constructions of type x. Effects of the sort observed here are then explained as being due to some of the more complex cases being tractable in terms of maximally simple procedures while others are not. Given the results described above, the principle seems to have a measure of plausibility as at least a working hypothesis.

The Comprehension of Focussed and Non-Focussed Pronouns

Jeanette K. Gundel and Deborah A. Dahl
University of Minnesota

Gundel (1980) has shown that focussed pronouns like the underlined form in (1) and non-focussed pronouns like the underlined form in (2) have different communicative functions.

(1) Q. Who did they call?
A. Pat said they called HER.

(2) Q. Has Pat been called yet?
A. Pat said they called her TWICE.

The purpose of this paper is to investigate the possibility that these different functions are associated with different psychological processes underlying the comprehension of pronouns.

We will begin with a brief description of the linguistic differences between focussed and non-focussed pronouns. The focussed pronoun in (1) is a referring expression. Its function, like that of other referring expressions, is to pick out and call the addressee's attention to, some entity in the discourse context. Thesediffer from full NP's only in that the entity is assumed to be identifiable on the basis of its presence in the immediate linguistic or non-linguistic context. As with full NP's, focus on a pronoun is obligatory when the pronoun is part of the comment of a sentence (i.e. new information being asserted, questioned, etc. about a topic), as in (1). Pronouns which are topics can also be focussed, e.g. if there is a topic shift or contrast as in (3)-(5).

(3) I asked Bruce about it. HE said he didn't CARE.

(4) THEM, I don't LIKE.

(5) Q. Are Bill and Mary still here?
A. HE went HOME, but SHE's in the other ROOM.

Non-focussed pronouns like the underlined form in (2) have no independent referring function. They are always controlled by already established discourse or sentence topics. From a communicative point of view, they are almost completely redundant. Their function is thus primarily syntactic. This distinction between focussed and non-focussed pronouns is independent of whether the coreferential full NP is in the same sentence or in a previous sentence in the discourse. Furthermore, as seen in (6) and (7), both focussed and non-focussed pronouns can be non-linguistically evoked.(Halliday and Hasan (1976) call this exophoric.)

(6) (A sees B reading an application and says)
Do you think we should ADMIT her? (non-focussed)

(7) (A hands an application to B and says)
Do you think we should admit HER?

Although focussed and non-focussed pronouns have different lexical, syntactic and semantic properties, linguistic theories of anaphora have not generally distinguished the two. This is no doubt due partly to the fact that most theories of pronominal anaphora are based on English, which has identical forms for focussed and non-focussed pronouns and where the latter differ from the former only in that they are unstressed and have corresponding phonological reduction in casual speech. The one exception is the 3rd person neuter singular it, which is always non-focussed.(see Linde (1979) .For example

(8) Q. Which do you want?
A.*I'll take IT.

(9)* IT, I don't LIKE.

However, in many languages non-focussed pronouns differ lexically from focussed ones. Some languages (e.g. Irish, Spanish and Polish) have long form focussed pronouns and corresponding short form , usually clitic, non-focussed pronouns. Compare the Polish examples in (10) and (11).

(10) Jan je tutaj. Ja go widziałam. (non-focussed)
is here I him saw
"Jan is here. I saw him."

(11) Q. Kogo widziałaś? "Who did you see?"
who saw

A. Ja JEGO widziałam. "I saw HIM" (focussed)
I him saw

In most languages, the two sets of pronouns (commonly referred to as non-emphatic and emphatic respectively) are related historically, the non-focussed pronoun being a phonologically reduced version of the focussed one. There are languages however which have totally unrelated forms for focussed and non-focussed pronouns. For example, in Fijian the 3rd person singular non-focussed form is e and the corresponding focussed form is koya. Finally, there are languages which allow, and in some cases require, so-called zero anaphora (i.e. no form at all) in those cases where English would have a non-focussed pronoun. The following Mandarin example from Li and Thompson (1979) is an illustration of this.

(12) qǔ-le shuǐ lái "(He) brought the water."
bring-aspect water

Focussed pronouns, however, cannot be omitted in these languages.

While the existence of zero NP-anaphora in languages like Spanish, for example, has often been linked to the fact that such languages have subject agreement marking on the verb, it is important to point out that such agreement is neither a necessary nor a sufficient condition for zero NP-anaphora. As can be seen from the example in (12), Mandarin allows zero NP-anaphora even though it has no agreement marking on the verb.

In addition to the lexical differences discussed above, focussed pronouns differ from non-focussed pronouns in their syntactic properties and in conditions on coreference with other NP's . In English, non-focussed pronouns are excluded from certain syntactic environments. For example, a direct object must precede an indirect object if the direct object is a non-focussed pronoun , as illustrated in (13) and (14).

(13) Q. Did you give the books to Tom?.
A. No. I gave MARY the books
*them

(14) Q. Which books did you give to Mary?
A. I gave Mary THEM.

As (13) shows, it is not non-focussed NP's in general, but only non-focussed pronouns, which are excluded from final position in such sentences.

On the other hand, some syntactic environments require non-focussed pronouns, as illustrated in (15) and (16).

(15) My pocket has a hole in it
*THAT

(16) The old dog still has a lot of life left in him
*HIM

Focussed pronouns also differ from non-focussed pronouns in conditions on coreference. For example, as noted in Akmajian and Jackendoff (1970) only non-focussed pronouns can be coreferential with a following full NP in the same sentence. Thus, John and he can be coreferential in (17) but not in (18).

(17) After he woke up, John went to TOWN.

(18) After HE woke up, John went to town.

Coreference assignments also differ depending on whether the pronoun is focussed or non-focussed in examples like (19) and (20).

(19) Mary called Alice and then SARAH called her.

(20) Mary called Alice and then Sarah called HER.

Since there are clear linguistic and functional differences between focussed and non-focussed pronouns, we would now like to address the issue of whether these differences are reflected in differences in processing.

We suggested above that focussed pronouns can refer to any entity in the immediate discourse context. Non-focussed pronouns, on the other hand, can only be coreferential with a subset of these, namely established topics. If this is true, one might expect that the processing of non-focussed pronouns would be less complex since the set of available entities is more restricted.

Theories of psychological processes underlying pronoun comprehension have generally not distinguished between focussed and non-focussed pronouns. Assumptions about pronoun processing can be classified into two main categories. The first, which we will refer to as the Reference Search Hypothesis is stated explicitly in Clark and Clark (1977, p. 78): "on finding a definite noun phrase, search memory for the entity it was meant to refer to and replace the interpretation of the noun phrase by a reference to the entity directly." A similar statement is found in Clark and Sengul (1979): "When listeners encounter 'the woman' or 'she' they are assumed to treat this as given information for which they must find a referent. They then search memory for the unique entity to which 'the woman' or 'she' was intended to refer."(Note that under this hypothesis pronoun comprehension is not assumed to be essentially different from comprehension of full NP's.)Caramazza and Gupta (1979) also seem to implicitly accept the Reference Search Hypothesis when, in discussing some of their stimulus sentences, they say:"the sentence materials used in Experiment I could be expected to generate this chain of events because the preposed subordinate clauses do not contain enough information to guide the subject in the search for appropriate referents to the anaphoric pronouns (emphasis added)

The second major kind of pronoun processing theory may be referred to as the Topic-Stability hypothesis. A pronoun, in this view, serves not as a signal to the listerner to initiate a memory search, but rather as a signal to assign coreference relations between the pronoun and the discourse topic. A statement consistent with this view, though not specifically proposing a processing theory, can be found in Chafe (1974): "if the explanation in terms of consciousness is correct, it is misleading to speak as if the addressee needs to perform some operation of recovery for given information. The point is rather that such information is already on stage in the mind." Karmiloff-Smith (1980) takes a similar view: "anaphoric pronominalization functions as an implicit instruction for the addressee not to recompute for retrieval of an antecedent referent, but rather to treat the pronoun as the default case for the thematic subject of a span of discourse." She goes on to say that deviations from the default (topic) case will be signalled by the use of a full NP. As we have seen (e.g. example (3) above) such deviations can be signalled as well by the use of a focussed pronoun.

If the two pronoun functions discussed above are associated with differences in processing, it may be that both the Reference Search Hypothesis and the Topic-Stability Hypothesis are correct. Reference search may be used in processing focussed pronouns and topic-stability may be used in processing non-focussed pronouns. Assuming that executing a reference search is a more complex task than assignment of reference to a discourse topic, this dual process hypothesis predicts that processing focussed pronouns should be more difficult than processing non-focussed pronouns. This hypothesis would also seem to predict that no differences should be found in processing between sentences which both have non-focussed pronouns. Previous work, however, has found such differences. Caramazza and Gupta (1979), for example, found differences in reaction time to naming the NP coreferential with a non-focussed (it would have been unstressed had it been presented auditorily) pronoun depending on pragmatic and syntactic factors. According to the Topic-Stability Hypothesis, these differences should not have been found. However, the prediction made by this hypothesis depends on the fact that unstressed pronouns must refer to the discourse topic. Sentences presented in isolation, as in Carmazza and Gupta's study, may be ambiguous with respect to discourse topic. If more than one entity is eligible to serve as the discourse topic, then a reference search might be necessary even for an unstressed pronoun, although the number of entities to be examined might be smaller than for a focussed pronoun. Thus, these results do not seem to provide a serious counterexample to our proposal.

The only previous study to directly address the question of comprehension differences between focussed and non-focussed pronouns is Maratsos (1973). He found that focussed pronouns of one type (that is the role-switch type seen in (20)), were more difficult for children to comprehend than unstressed pronouns. He interprets this difference to the operation of a role stability strategy that tells the child to try to maintain the same actors in syntactic and semantic roles. Obviously this interpretation resembles the Topic-Stability hypothesis, in that in the Topic-Stability hypothesis the listener is maintaining the same entity as a discourse topic.

In the dual-process hypothesis, as well as in Maratsos's hypothesis, stress is seen as signalling change, however, the fact that some types of focussed pronouns do not signal grammatical or semantic role change (such as the example in (7)) shows that Maratsos's characterization is not quite accurate for a wider sample of focussed pronouns than the ones he used in his experiment.

In light of the results that Maratsos obtained, we expected that the processing of focussed pronouns, assumed to be of the reference search type, would be more difficult than the processing of unfocussed pronouns, assumed to be of the topic-stability type.

We conducted a pilot study to test the hypothesis that there are different kinds of processing for stressed and unstressed pronouns. In this study, 15 subjects listened to a set of 40 short discourses, 20 experimental and 20 filler. The experimental discourses occurred in two identical forms, except with a biasing context that made a focussed or non-focussed pronoun appropriate. In both forms of the discourse the referent of the pronoun was the same. For example:

(21) A. Did Bill say who would be late?
B. Yes, after I called him up, Bill said that HE would be late.

(22) A. Did Bill say whether he would be late?
B. Yes, after I called him up, Bill said that he would be late.

After listening to the sentences, the subjects answered true or false to a statement about a part of the sentence that had nothing to do with the pronominal reference. After (21) or (22), for example, the subjects might hear:

(23) True or false: Bill called me up.

On the assumption that a difficulty in pronoun processing would lead to a general degradation in performance in comprehending all parts of the sentence, we predicted that more errors would be made in the sentences with focussed pronouns in them.

We found that subjects did not made significantly more errors in either condition. We believe, however, that this result is due to the fact that the task issimply too easy. We found that subjects performed at about the 90% level of correctness for both focussed and non-focussed pronoun discourses, and most of these errors were due to two stimulus discourses that were difficult for other reasons. We also asked the subjects to judge whether they felt the following statement was easy, medium, or difficult to answer, and found no difference between focussed and non-focussed conditions in this measurement either. We did find that judgements of difficulty were not consistently related to performance. Very often subjects thought that the statement was easy to confirm when they gave the wrong answer.

A future experiment will provide a more sensitive test of this hypothesis. This experiment will auditorily present subjects with discourses containing pronouns of the types we are interested in, and measure reaction time to naming the referent of the pronoun. This would be similar to the procedure used in Caramazza and Gupta (1979), except that they used a visual presentation.

If our hypothesis is correct, then reaction time to naming the referent of a focussed pronoun should be longer than reaction time to naming the referent of a non-focussed pronoun. If there is no difference in reaction times, we would conclude that the reference search process is used for all pronoun processing. Note that the Topic-Stability hypothesis cannot be correct (at least for adults), for all pronouns, since it can lead to the wrong referent's being selected for some focussed pronouns, as illustrated in (19) and (20). On the other hand, reference search could lead to the correct referent for all pronouns.

We have suggested that pronouns can be divided into two types on the basis of their linguistic characteristics, and we have suggested that people may be able to take advantage of the severe restrictions on what the antecedent of an unstressed pronoun can be in processing. No differences were detected in a pilot study, but an additional experiment is proposed that will use a more sensitive test of processing difficulty than the one used in the first experiment.

References

Akmajian, A., and Jackendoff, R. (1970). Coreferentiality and stress. Linguistic Inquiry, I, 124-126.

Caramazza, A., and Gupta, S. (1979). The roles of topicalization, parallel function, and verb semantics in the interpretation of pronouns. Linguistics, 17, 497-518.

Chafe, W.L. (1974). Language and consciousness. Language, 50, 111-133.

Clark, H.H., and Clark, E.V. (1977). Psychology of Language. New York: Harcourt Brace Jovanovich, Inc.

Clark, H.H., and Sengul, C.J. (1979). In search of referents for nouns and pronouns. Memory and Cognition, 7, 35-41.

Gundel, J. (1980). Zero-NP anaphora in Russian: A case of topic-prominence. In Kreiman, J., and Ojeda, A.E., eds., Papers from the Parasession on Pronouns and Anaphora. Chicago: Chicago Linguistic Society, 139-146.

Halliday, M.A.K., and Hasan, R. (1976). Cohesion in English. London: Longman.

Karmiloff-Smith, A. (1980). Psychological processes underlying pronominalization and non-pronominalization in chidren's connected discourse. In Kreiman, J., and Ojeda, A.E., eds., Papers from the Parasession on Pronouns and Anaphora. Chicago: Chicago Linguistic Society, 231-250.

Li, C.N., and Thompson, S.A. (1979). Third-person pronouns and zero-anaphora in Chinese discourse. In Givon, T., ed., Syntax and Semantics 12: Discourse and Syntax. New York: Academic Press, 311-335.

Linde, C. Focus of attention and the choice of pronouns in discourse. (1979). In Givon, T., ed., Syntax and Semantics 12: Discourse and Syntax. New York: Academic Press, 337-354.

Maratsos, M.P. The effects of stress on the understanding of pronominal co-reference in children. Journal of Psycholinguistic Research. 2 (1973), 1-8.

The Natural Natural Language Understander

Henry Hamburger, UCI and NSF
and
Stephen Crain, U of Texas

This study of natural language comprehension by natural understanding systems (children) is based on a procedural analysis represented in the form of a programming language. To clarify what is cognitively required for a child to respond appropriately to certain expressions in English, we show how these forms can be translated into procedures in a high-level programming language. It is then possible to discuss two kinds of difficulties a natural language form can present to the listener: (i) incompatibility of the form with its associated procedure and (ii) complexity of that procedure. An example of procedure complexity is the nesting of loops, whereas a contributor to incompatibility is a word or contiguous phrase that corresponds to separated pieces of the procedure. We shall present evidence of both types of difficulty from experiments with children and will compare the predictions of our procedural view with those of a less detailed syntactic explanation that has been advanced for a subset of the phenomena.

Many cognitive tasks can be expressed either as a natural language command or as a programming language procedure. In many tasks requiring some elementary mathematical knowledge (e.g., counting), the cognitive procedure appears to be substantially more complex than the syntactic structure of the command. In such tasks, an explanation of children's difficulties can be pursued more profitably in the realm of cognitive procedure than of syntactic structure or parsing.

To take a particularly simple example, the verb 'count' has the capacity to serve as either a transitive or an intransitve verb, but it is probably never the first such verb encountered by a child; compare 'let's eat' and 'eat your cracker.' Therefore the transition from intransitive use of 'count' to its transitive use involves no syntactic innovation for a child. But despite being syntactically ordinary, 'count' presents complexities both in its procedure and in translation to that procedure. Just ask the next three-year-old you see to count, and then to count a few objects. Chances are good that you will get errors on the transitive (latter) task but not on the intransitive one. A look at these errors will give some perspective on what a correct procedure must do.

One kind of attempt at transitive counting involves blithely swinging a finger past the objects to be counted, while apparently counting essentially intransitively, with no attempt to coordinate individual objects to individual numbers. A somewhat more sophisticated performance has the finger stopping at some or all of the objects, but in imperfect coordination with the numbers. These misperformances, we believe, arise and persist not out of ignorance of the lexical item 'count' or the associated data structure of positive integers, nor from the syntax of transitive verbs (which, it was suggested above, has already been learned), but because of difficulty inherent in forming a correct procedure for the task of transitive counting.

Consider procedures i and ii, intended to represent correct procedures for the two kinds of counting, expressed at a coarse enough grain that individual statements can be taken as having theoretical content. The tokens are intended to be more or less self-explanatory. In the interest of simplicity, no mention is made of initialization, stored data, or output.

```
i  procedure:  'Count.'
     repeat                 The procedure runs
       next number          through successive numbers
     until fail             until there are no more.

ii procedure: 'Count them.'
     repeat
       next object          At each cycle, the next
       next number          object and number are
     until fail             noted.
```

For i to be utilized in forming ii, it is necessary to insert new material, specifically 'next object,' into the loop already present in i. Thus one approach to accounting for the difficulty posed by the transitive case is to hypothesize that breaking into a loop is a difficult or low-priority move for acquisition or comprehension. A related view is implicit in iii where it can be seen that a single word, 'count,' corresponds to two separate pieces of procedure. This phenomenon can appropriately be called translational discontinuity in the sense that it pertains to the relationship betweeen two representations. Since it arises in the translation from a (very) high-level language (English) to a less-high-level language (of procedures), it can also be called a compiling discontinuity. Note that there is no discontinuity in the phrase structure of the language form itself.

```
iii procedure: 'Count ...'
      repeat
        ...
        next number
      until fail
```

To take these ideas a step further, consider the phrase 'the second green ball.' Roeper (1972) found that many children interpret this phrase as if 'second' and 'green' each modify the noun in the same way, that is, as if the phrase meant something like 'the ball which has the properties of being both green and second.' Matthei (1978) expresses this observation as a distinction in syntactic phrase structure, assigning the adult interpretation the phrase structure '(second (green ball))', and the non-adult interpretation '(second green ball).' The idea of using syntactic phrase structure in this way to encode semantic interpretations must rely on some assumptions about how semantics relates to syntactic structure, for example, a compositional semantics tied to the syntactic phrase marker. In any case, no such phrase-structural distinction is possible for phrases like 'second ball,' for which Matthei found substantial corresponding errors: 36% of responses (four- and five-year-olds, mostly) interpreted 'second ball' as 'the one that is both second and a ball,' as opposed to the 52% who made the similar error on the example above that had an adjective in the phrase.

Pursuing the procedural approach with this construction, we now find not only compiling discontinuity but also nested loops. These aspects of the procedure appear both with and without the adjective, as can be seen with reference to iv and v, again omitting initialization and any data structures.

```
iv procedure: 'second ball'
     repeat
       repeat
         next object
       until
         pred ball
       next number
     until
       pred two

v  procedure: 'second green ball'
     repeat
       repeat
         next object
       until
         pred green
         pred ball
       next number
     until
       pred two
```

The task environment is a display of several objects in a row, with a left-to-right ordering clearly communicated in advance. The child is asked to 'take the second ball' from a display of balls and boxes, or to 'take the second green ball' from a display of red and green balls. To take the former case, suppose that the first object is a box. Then even if the second object is a ball, so that it is both second and a ball, it is still not the second ball. Procedure iv correctly makes this distinction (which eludes so many children) by means of nested loops in which successive objects are checked for ballhood in the inner loop and counting proceeds in the outer only in case the current object is a ball.

With an adjective present the appropriate procedure is v, which is like iv except that the inner loop is exited only in the event that two successive predicates are satisfied. Rather than introduce an 'and' operator in v, we have allowed the possibility of multiple exit tests, thereby making conjunction look simpler than disjunction. Even so, one still should expect the extra predicate to add some processing burden and indeed the phrase in v does lead to more errors than that in iv. Furthermore the outer predicate should also impose a processing burden. Suppose we remove it and interpret the absence of any exit test as signifying 'repeat forever or until system breakdown (say by failure of a 'next' operator).' What is then left of v is a procedure for 'count the green balls,' which does indeed appear to cause children less difficulty than 'second green ball.' (Matthei regarded the two as cognitively equivalent).

Not only do the words 'second' and 'ball' together translate into nested loops, but they do so in a discontinuous manner. This is so because 'second' is responsible for the outer loop and 'ball' (in iv) is responsible for the inner loop. Worse yet, the programming statement 'next object' is implicit in the counting loop, so that under an absolutely left-to-right control structure, the state of (cognitive) affairs after 'second' is processed would be as in vi. Processing the remainder of the phrase requires, on this model, locating 'next object' in the existing loop, using it ('next object') in constructing a new loop, and finally nesting the new loop in the existing one.

```
vi procedure: 'second ...'
     repeat
       ...
         next object
       ...
         ...
       next number
     until
       pred two
```

It is our view that these complexities are what make such phrases difficult for children. Not only does this account indicate where complexities lie, but in addition it provides an account of how specific errors can arise from a straightforward attempt to avoid breaking into loops and incurring compiling discontinuity. Suppose that a child simply tacks on the appropriate test ('pred ball') at the end of vi to form vii. This is possible if the convention introduced in v, above, is used again here: allowing a sequence of 'pred' statements as a compound exit test. This ploy yields precisely the observed incorrect result whenever it is possible (when the second object is a ball), and an infinite loop otherwise.

```
vii procedure: 'second ball'
                (incorrect interpretation)
      repeat
        next object
        next number
      until
        pred two
        pred ball
```

To gain perspective on the earlier examples and raise some new issues, consider the phrase 'second biggest ball.' Here the ordering along which 'second' is to be counted is based not on position but on size. This ordering must be at least partially determined by the subject, using an appropriate algorithm. The usual sorting algorithms (ripple, bubble, Shell, etc.) are probably poor models both because they are severely nonparallel and because they provide a complete ordering where only a partial one is needed.

In addition to some sorting, this task demands memory of some order relationships that have been established by that sorting. These size relationships in short-term memory must be used in concert with the ordering of positive integers held in long-term memory. This requirement is not present for the earlier tasks since the positional ordering is continually available from the display. In our experiments the display has been a card with a row of pictured objects, so subjects cannot create a physical order. If separate physical objects were used, then a child's sequence of moves might reveal use of a specific sorting algorithm.

Results of pilot experiments suggest that 'second biggest ball' is, as one would expect from these considerations, an extremely difficult phrase to interpret. We devised the most straightforward we could that would test comprehension of this phrase: all the objects were identical except in size; all but two were of the same small size; the remaining two were both substantially larger than the smaller ones, adjacent to each other and noticeably different from each other in size; neither was in second position. In experiments so far the error rate has been 80%; we will report more comprehensive testing at the conference.

It is possible to set up more complex displays in which different miscomprehensions lead to distinct choices as response. With balls and boxes of various sizes there exist arrays in which, say, one object is both second biggest and a ball but is not the second biggest ball; or in which the second ball is the biggest; etc. A welter of possibilities exists for testing the way in which ordinals, superlatives, relative adjectives, absolute adjectives, nouns and relative clauses are comprehended by children and what the course of development looks like, in terms of the kinds of procedures we have posited here. Such developmental sequences are, in turn, the raw material for theories of an acquisition device.

Why Do Children Say "Goed"?
A Computer Model of Child Generation

Mallory Selfridge
Department of EE and CS
University of Connecticut, Storrs, CT

1.0 Introduction

An important question in modelling child language generation is why children say regular forms of irregular words, such as "goed", (Clark and Clark, 1978) during development, although they never hear them. Three other general characteristics of children's generation also require explanation. First, Benedict's (1976) work suggests clearly that comprehension of various aspects of language precede the generation of those aspects. Second, the length of the utterances children say becomes generally longer as development proceeds (Bloom, 1973). Third, Wetstone and Friedlander (1973) suggest that first children say things in the wrong order, and then say things in the correct order.

In order to address these issues, this paper explores the hypothesis that learning to talk is driven by learning to understand. This hypothesis begins by assuming that the principal effect of learning to understand is the development of the lexicon as additional words are learned and their "definitions" are refined and modified. It further assumes that the language generation process is not learned, but is an innate part of a child's cognitive repertoire. Finally, it states that the ability to generate grows as the lexicon develops during the development of comprehension. The hypothesis predicts that a computer model which incorporated it would display the characteristics described above.

CHILD (Selfridge, 1980) was initially developed to model a subset of the development of comprehension about a limited domain in a child between the ages of one and two years, using context-based inference and learning rules to build a dictionary accessable to a conceptual analyzer (Birnbaum and Selfridge, 1979). New words were added to the dictionary, the meanings of ones it had learned were refined, and syntactic information on how to combine word meanings were stored under appropriate words. Meaning was modelled using Conceptual Dependency (Schank, 1973), while syntactic knowledge was represented using Sequential Structure (Selfridge, 1980) which encodes the utterance position of the filler of a slot of a word meaning. Thus CHILD's learning was based upon meaning, with syntactic knowledge indexed upon this meaning.

CHILD has now been equipped with a generator (Cullingford, Krueger, and Selfridge, 1981), which has access to the dictionary built up during comprehension learning. CHILD now learns to generate in the same limited domain, and offers explanations for the psychological phenomena described. In particular, although CHILD does not yet say "goed", it provides a precise account of how it could be given experiences which would lead it to say "goed".

2.0 Learning to Generate

Children speak in many different situations. CHILD only generates language in one of these, that in which a child describes something observed (Halliday, 1975). The user teaches CHILD to understand by providing utterances in situations in which CHILD can infer their meaning. To demonstrate the development of generation, the user provides CHILD with a Conceptual Dependency concept, simulating visual input. When given such a concept, CHILD attempts to generate it.

The following sequence of utterances interspersed with transitions summarizes the development of CHILD's generation capacity, and is drawn from a full run of CHILD during which it both learns to comprehend and generate. The statements referring to CHILD "learning" meaning and syntax summarize the learning of comprehension described in Selfridge (1980). In order to show development, CHILD has been supplied with the same concept to generate repeatedly: the concept for "the parent puts the ball on the table", (PTRANS ACTOR (PARENT1) OBJECT (BALL1) TO (TOP VAL (TABLE1))).

CHILD knows no words

CHILD says "um"

CHILD learns meaning of "ball"

CHILD says "ball"

CHILD learns meaning of "put"

CHILD says "ball put"

CHILD learns meaning of "table"

CHILD says "table ball put"

CHILD learns syntax of "put"

CHILD says "put ball table"

CHILD learns meaning of "on"

CHILD says "put ball table on"

CHILD learns syntax of "on"

CHILD says "put ball on table"

This progression shows that CHILD's generation does correspond to the data described earlier. First, CHILD certainly does learn to generate after it learns to understand. Second, the length of its utterances does indeed grow. Third, the ability to generate structurally correct utterances follows the ability to generate incorrect utterances.

3.0 Why Would CHILD Say "Goed"?

There are several properties of CHILD which would lead it to say "goed". First, its dictionary is ordered and the first word found during lookup which matches the concept being generated is used. Second, words are not forgotton and removed from the dictionary, but they may be "covered-up" by words learned later. Third, both new words and words whose meaning has been modified are placed at the top of the dictionary where they will be found first during lookup. The following developmental progression accounts for why children say "goed" according to the CHILD model. Each of the proposed learning events can be modelled by CHILD's existing learning rules.

The first stage is that in which children learn "go" and "went" as meaning PTRANS with the TIME slot containing PRESENT and PAST respectivly. The order in the dictionary is, top to bottom, "went" and "go". "Went" is on the top because it is presumably learned later than "go." At this stage the child will use these words correctly, since when he wants to generate a PTRANS with TIME as PAST he will lookup "went" directly, and when he wants to generate PTRANS with TIME as PRESENT, he will find "go" directly.

The second stage occurs when he learns "ed" as a separate lexical item, whose meaning is the filler PAST. This is learned during comprehension in the

same way other words are learned, and as a result "ed" is placed on the dictionary, whose order is now "ed", "went", "go". Although "ed" is now in the dictionary, the child desiring to generate PTRANS with TIME either PAST or PRESENT will still use "go" and "went", since dictionary lookup matches on the root concept, and the meaning of "ed" doesn't match PTRANS.

At the third stage, the child learn that the meaning of "go" does not include the PRESENT filler of the time slot, perhaps because the child is learning about "go" and the future tense. He must also learn that the filler of the TIME slot of "go" must follow "go", perhaps because he learns to understand "go ing". Learning a modified meaning of "go" results in "go" being placed on the front of the dictionary, and its order is now "go", "ed", "went". At this stage, the child desiring to generate PTRANS with TIME as PRESENT will find the meaning of "go" matching this PTRANS, and will then search the dictionary to generate the PRESENT sub-concept. Since there is nothing there, he will use "go" alone. However, to generate PTRANS with TIME as PAST, he will again find "go" matching the PTRANS, and will search the dictionary to express the PAST subconcept. This time, however, he will find "ed", and will thus generate "goed".

The fourth stage occurs when the child hears "went" again, perhaps in ordinary discourse or as a parental correction to "goed". This correction results in "went" being placed in the front of the dictionary, whose order is now "went", "go", "ed". At this point, the child will use "go" for PTRANS with TIME as PRESENT, "ed" is available for expressing the PAST time for ordinary action words, but he will cease using "goed", since he finds "went" expresses PTRANS with TIME as PAST directy.

4.0 Conclusions

CHILD offers an explanation for the psychological data described earlier. Generation follows comprehension because generation cannot occur until comprehension learning adds words to the dictionary. Length of utterances increases because the number of words available to express a concept increases during comprehension learning. Utterances are generated with incorrect structure before they are generated with proper structure because syntax is indexed under word meanings, word meanings are learned before their syntax, and word meanings without syntax are available to the generator before word meanings with syntax. In particular, this paper proposes a specific explanation of why children say "goed": because generation is driven by understanding, because the dictionary is ordered and the first word found matching the concept to be generated is used, and because new words and words with refined meanings are placed on the top of the dictionary.

This account of why children say "goed" makes four specific predictions. First, since "went" is still in the dictionary even though the child says "goed", the child will still understand the utterances containing "went" during the stage at which he is generating "goed". Second, since "went" was relearned as a result of an experience specifically with "went", this model predicts that a child will continue to say "goed" at least until he has heard "went" again; that is, after he begins to use "goed", he will never say "went" before he hears "went" again. Third, since corrections or relearning are specific to individual words, the transition from the third to fourth stages for various irregular words will occur individually. That is, if the child is saying "goed" and "runned", learning "went" again will not result in the child also saying "ran". Rather, a specific experience with "ran" is needed. Fourth, this model predicts that a child will never say "goed" until after he has learned to understand "go" in a tense other than the present, because this experience teaches him that the TIME slot under "go" is empty, as is needed to say "goed".

Acknowledgements

Dr. Roger Schank's assistance in this work was invaluable. Dr. Richard Cullingford has made many valuable suggestions, and Dr. Katherine Nelson has provided many insights into problems of modelling child language learning. Larry Birnbaum, Marie Bienkowski and Jamie Callan have contributed both ideas and criticisms.

Bibliography

Benedict, H. (1976). Language Comprehension in 10 to 16 Month-old Infants. (thesis) Department of Psychology, Yale University, New Haven, Connecticut.

Birnbaum, L., and Selfridge, M. (1981). in Inside Artificial Intelligence: Five Programs plus Miniatures. Schank R. and Riesbeck C.K. (eds.), Lawrence Earlbaum Associates, Hillsdale, NJ.

Bloom, L., 1973. One Word at a Time. Mouton, The Hague.

Bruner, J.S., Goodnow, J. J., and Austin, G.A., (1956). A Study of Thinking. John Wiley and Sons, New York

Clark, E.V. and Clark, H.H., (1978). Psychology and Language, Harcourt Brace Jovanovich.

Cullingford, R.E., Krueger, M.W., and Selfridge, M. (1981) A Picture and a Thousand Words: Automated Explanations as a Component of a Computer-Aided Design System. Forthcoming.

Halliday, M.A.K., (1975). Learning How to Mean: Explorations in the Development of Language Edward Arnold, London.

Schank, R. C., (1973). Identification of Conceptualizations Underlying Natural Language. In R. C. Schank and K. M. Colby (eds.) Computer Models of Thought and Language W.H. Freeman and Co., San Fransisco.

Selfridge, M. (1980). A Process Model of Language Acquisition. Computer Science Technical Report 172, Yale University, New Haven, Ct.

Westone, H. and Friedlander, (1973). The Effect of Word Order on Young Children's Responses to Simple Questions and Commands. Child Dev. 44:734-740

Winston, P. (1975). Learning Structural Descriptions from Examples. In P.H. Winston (ed.) The Psychology of Computer Vision. McGraw-Hill, New York.

FIVE EXPERIMENTS IN THE DEVELOPMENT OF THE EARLY INFANT OBJECT CONCEPT

BY

George F. Luger,
University of New Mexico
Department of Computer Science
&
T.G.R. Brower & Jennifer G. Wishart,
University of Edinburgh
Department of Psychology

Abstract

A computational model is proposed for the early stages of development of the object concept in infants. Stages in development are represented as a sequence of grammars or rewrite rules parsing a set of perceptual phenomena. The infants' changes between developmental stages can be described by differences between the grammar rules modelling each stage. The program replicates five Bower et al studies on the development of the object concept and reaffirms the primacy of rest and motion parameters as explanatory invariants in early object concept development.

Forty years ago, Piaget first observed and described the problems infants have in understanding the nature of objects (Piaget, 1936, 1937, 1946). Piaget identified six invariant stages of development through which the infant comes to cope with and understand external objects.

One alternative theory of object concept development which has been proposed is the identity theory of Bower and Wishart (Bower, 1967, 1972, 1974; Wishart, 1979). 'Identity theory' believes the infant's problem in understanding the nature of objects to lie in discovering the spatio-temporal relationships which underpin the identity of any object. While Piaget believed the young infant to have no true differentiated awareness of either self or objects in the first year of life, Bower and Wishart believe a basic understanding of the independent existence of both to be present at a very early stage, possibly even at birth. The infant is seen as having difficulty, not in differentiating himself and his actions from other activity in the external world, but rather in understanding the unique nature of objects themselves.

The identity hypothesis can be summarised in the form of a series of three conceptual rules (Bower, 1974, 1981; Wishart, 1979) which can cover all six stages of search behavior found in standard Piagetian object permanence testing situations and can also explain the otherwise puzzling behavior shown by infants in other object-related tasks. (See e.g. Bower & Wishart, 1973, Butterworth, 1977, Wishart & Bower, 1981). While it is the eventual aim of our research to develop a program to cover the entire sequence of object concept development, this paper will report only our initial work on modelling Stages I - II of that development, a period which, according to Bower and Wishart, sees the development of the first two conceptual rules for attributing identity to successive appearances of any object.

These rules can be formulated as a sequence of grammars or rewrite rules that translate perceptual phenomena into sets of behavioral responses. The perceptual phenomena will remain constant throughout the period of testing, i.e., across the running of the sequence of grammars that represents the early stages of development.

This paper hypothesizes that the symbolic output of the featural and motion detection mechanisms is available to the cognizing subject. This model offers no explanation of the physiological origins of such phenomena, but rather emphasizes the descriptive adequacy of the internal symbol structures and the interpretive adequacy of the cognizing subjects' manipulation of such symbol structures. Further, the changes in the computational rules expressing the interpretive adequacy of infants at various stages in development will offer explanation of that development.

Each experiment of this study is composed of a sequence of "snapshots" representing the physical situation according to a time parameter. Snapshots represent objects by themselves, partially or totally obscured by occluders, and replaced by other objects. An example of such a set of snapshots may be seen in Figure 1. This figure represents a subset of the snapshots taken from Experiment 5 of Section II, where S(n) indicates the time parameter.

A set of symbols represents each snapshot and a set of rules characterizes the grammar

that interprets each sequence of snapshots Each rule of a grammar represents a different interpretive capacity of the subject such as to locate an object symbol structure within a fixed radius r of a spatial position (x,y,z).

The grammars of this model are designed to implement the rules outlined by Wishart (1979). An implicit assumption is that the infant is motivated to maintain contact throughout the event sequence with the object he/she initially identifies as interesting.

This computational description of the object permanence phenomenon is written in PROLOG (Warren and Periera, 1977; Warren, 1979). The action of PROLOG is of a unification algorithm operating on a set of record structures. These structures are of two general forms: a set of facts and a set of inference rules. The PROLOG facts are used to make-up the object structure for the description of each snapshot. For example the facts loc (OBJN, X, Y, Z, T), color (OBJN, CL T), and size (OBJN, SZ, T) indicate OBJECTN has color CL, size SZ and location (X,Y,Z) at time T. The combination of these and other descriptors make up each snapshot of the experiment. The grammar rules interpret these object structures.

PROLOG rules are of the form "A :- B,C,D," which may be described procedurally as "to accomplish A attempt to accomplish B and C and D." B, C, and D may be facts (checked to be true) or may themselves be rules that lead to the proof or performance of B, C, and D. For example, Grammar 1 says to test for a permanent* object at a location look for:

1) An object structure within a fixed radius r of the location.
2) Check if previous snapshots would indicate a permanent object should be at this location.
3) Test the object structure for interest (parallax).
4) Check whether the object structure is intact (that is whether it or any of its boundaries are occluded), and finally,
5) Based on this object structure look to an appropriate position for the next object structure.

The grammar of the second developmental stage is similar to the grammar for stage one. The essential difference is a rule called twice by the grammar for the object structure of each snapshot. This rule represents the older infants ability to check different perceptual values such as size and color between snapshots. (Luger et al, 1981).

II. Five Experiments

In this section five experiments will be considered and the action of the computer model in these situations described. The model is in two parts, each representing a stage of development of the infant. The first part, Grammar 1, represents the earliest stage of development modelling Rule 1; Grammar 2 models Rule 2 (Wishart, 1979).

In the conclusion, the results of the computational model run on the experimental situations will be compared with the results of infants tested in the same situations.

*Although this paper does not elaborate on the distinctions between object permanence and object identity interpretations of object concept behaviors, the authors' bias towards identity theory should be made clear. A basic notion of object permanence (i.e. of the continued existence of objects when unperceived) is assumed to be present from very early on (Bower, 1967), according to identity theory, it is the step-like discovery of the precise spatio-temporal nature of that existence which lies behind the sequence of errors found in standard object concept tasks, not the development of a notion of permanence per se.

A. The Experiments

Experiment 1. (Bower, Broughton & Moore, 1971; Bower & Paterson, 1973).

An object, interesting to an infant, is moving on a straight path in full view of the infant. The object then stops in full view.

The young infant (Stages I II) typically pauses with the object for almost half a second and then continues to follow the objects' former path across the field of view (movement error). The older (Stage III) infant pauses with the stopped object, looks onto the path the object no longer follows, and then goes back tothe stopped object.

Experiment 2. (Bower, Broughton & Moore, 1971; Bower & Paterson, 1973).

An object is at rest in full view of the infant. It is displaced and follows a fixed path at constant velocity.

The younger infant follows the displaced object but then looks back to the position of the original stationary object (place error). The older infant pauses and then follows the path of the displaced object.

Experiment 3. (Bower, 1974)

An object moves along a path. It is replaced in mid-path by another object that continues to move along the path.

The young infant is not disturbed by the mid-path substitution. The older infants are confused moving their eyes back and forth between places of different objects.

Experiment 4. (Bower, Broughton & Moore, 1971)

An object moves across the field of view and behind a occluder. A different object reappears on the other side of the occluder.

The young infant follows the path of the object and continues to track the changed object when it appears. The older infant reacts to the new object when it reappears and looks back to the occluder.

Experiment 5. (Bower & Paterson, 1973)

An object moves across the field of view. It stops on one side for a brief period of time and then moves back in the opposite direction stopping in its original position again for a brief time. After repeating this sequence of moves four times it goes off in the opposite direction. See snapshots of Experiment 5 in Figure 1.

The younger infant continues to have movement and place errors as the object moves and stops and moves off again. With repeated exposure to the sequence, she may adopt a place to place strategy of jumping to the object's next stopping point as soon as it begins to move off from its present position. This strategy is only briefly successful since it cannot cope with the final change in direction. The older infant is not disturbed by the movement and stopping in either direction.

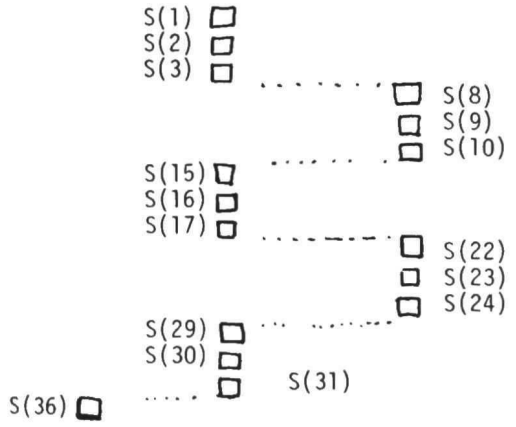


Figure 1. Snapshots from the motion and rest pattern that makes up Experiment 5.

B. The Results

Experiment 1.

Grammar 1 identifies an interesting object, say OBJ1, and associates this object with a path across the field of view. When OBJ1 stops moving Grammar 1 continues to follow the former path, not finding an object it looks back to the now stationary object, calls it OBJ2, and looks back searching again for OBJ1 on its original path. Grammar 2 in this situation identifies OBJ1 as interesting and follows it. When it stops, Grammar 2 continues on path and then looks back to the stopped object, notices its' similarity in size and shape to the original OBJ1 and continues to look at it.

Experiment 2.

Grammar 1 identifies an interesting object, OBJ1, stationary before it. When the object starts to move, Grammar 1 follows it, looks back to its former location and then looks to the new moving object and calls it OBJ2. Grammar 2 does the same as Grammar 1 except that it recognizes the size and shape of the moving object to be the same as that of the original stationary object, and does not create a new object.

Experiment 3.

Grammar 1 identifies OBJ1 and follows its path ignoring the mid-path substitution. Grammar 2 follows OBJ1 as interesting and noting the perceptual changes of the sustituted object calls this new object OBJ2 and continues to follow it.

Experiment 4.

Grammar 1 sees OBJ1 as interesting and follows it along. It continues to follow the path behind the occluder and follows the path as the new object emerges. Grammar 1 sees only one bounded volume of space moving on a path. Grammar 2, noting peceptual changes sees 3 objects; the original, the occluder and the new emerged object.

Experiment 5.

With every movement and pause Grammar 1 sees a different object for each movement and each place. Grammar 2 sees only one object, the perceptual checks at each motion translation allowing transition to new motion sequences without confusion.

III. Summary and Conclusions

Several researchers have recently reported on computational models of preverbal children's problem solving skills (Prazdny, 1980; Young, 1976). The behaviors modelled in these studies are as simple as eye movements, reaching, and gross body movements. Modern research techniques such as video tapes, however, allow researchers to be quite comfortable with the interpretation of such motor behavior. Such data makes up the Bower and Wishart studies, and this is the data our program interprets.

Grammars 1 and 2 follow fairly closely the two rules suggested by Wishart (1979). Grammar 2 with its perceptual checking for size, shape and color factors allows proper detection of changes between the two objects when these change,

and proper positing of no change when only the motion and rest factors differ. Thus motion and rest are established in the "identity" model as explanatory invariants of the first stages of infants' acquisition of the object concept.

In both experiments 1 and 2, Grammar 1 identifies 2 objects, one stationary and the second in motion. Grammar 2 with more advanced perceptual checking, notes the similarity of the object in motion and the object at rest and does not posit a second object.

In experiments 3 and 4 Grammar 1 sees only one object following a fixed path. Grammar 2, noting perceptual differences, posits new different objects for each perceptual change in the objects.

In experiment 5, Grammar 1 posits a new object for each path of movement and each stationary object position. Grammar 2 sees only one object.

An important area for continued work is to develop a third grammar capable of interpreting object structures in more complex situations than those dealt with above, in experimental situations, for example, involving partial occlusion of an object or close spatial interaction between a number of objects. Bower and Wishart have posited a third conceptual rule (Rule III) to account for these final stages of the development of object understanding (Piaget's Stage VI). The rule states that two or more objects cannot be in the same place or on the same path of movement at the same time unless they bear a spatial relationship to each other which involves the sharing of common boundaries. The interpretative adequacy of this rule could be evaluated against experimental situations in which such boundary sharing occurs, as, for instance, when an object is placed on top of a platform, inside a cup or behind a screen (Wishart & Bower, 1981). Experiments of this form make up a valuable part of the research to explore the object concept and will be the focus of the next stage of this research.

The ultimate aim of the research will, of course, be to model the rules which produce the change upwards from one conceptual stage to the next. A series of cost-gain acceleration studies with infants is at present in progress in an attempt to produce data which will give us some insight into these mechanisms for change (Bower, 1981).

Acknowledgements

The authors would like to thank the British Medical Research Council for their generous support of this research (Research Grant No. G979/932/N).

References

Bower, T.G.R., 1967, "The Development of Object Permanence," Perception and Psychophysics, 2, 411-418.

Bower, T.G.R., 1972, "Object Perception in Infants," Perception, 1, 15-30.

Bower, T.G.R., 1974, Development of Infancy, San Francisco, W.H. Freeman.

Bower, T.G.R., 1977, A Primer of Infant Development, San Francisco, W.H. Freeman.

Bower, T.G.R., 1981, Development in Infancy (2nd Edition), San Francisco, W.H. Freeman.

Bower, T.G.R., Broughton, J.M., & Moore, M.K., 1971, "Development of the object concept as manifested in changes in the tracking behavior in infants between 7 and 20 weeks," J. Exp. Child Psych., 11, 182-193.

Bower, T.G.R., & Wishart, J.G., 1973, "The effects of motor-skill on object permanence,: Cognition, I, 165-171.

Butterworth, G., "Object disappearance and error in Piaget's Stage IV task," J. Exp. Child Psych., 23, 391-335.

Luger, Bower & Wishart, "A Computational Model of Develoomment of the Object Concept in Infants," Department of Computer Science Technical Report, University of New Mexico, 1981.

Piaget, J., 1936, The Origins of Intelligence in Children, London Routilege and Kegan Paul, 1953. (Original French edition, 1936).

Piaget, J., 1937, The Construction of Reality in the Child, London, Routilege and Kegan Paul, 1954. (Original French edition, 1937).

Piaget, J., 1946, Play, Dreams and Imitation in Childhood, New York, Norton, 1951. (Original French edition, 1946).

Prazdny, S., 1980, "A Computational Study of a Period of Infant Object-Concept Development," Perception, 9, 125-150.

Warren, D. & Periera, 1977, "PROLOG - The Language and its Implementation Compared with LISP," ACM, SIGPLAN Notices, 12(8), and SIGART Newsletter, No. 64, 109-115.

Warren, D., 1979, "PROLOG and the DEC System 10," in (Michie, ed), Expert Systems in the Micro Electronic Age, Edinburgh, The University Press.

Wishart, J.G., 1979, The Development of the Object Concept in Infancy, Unpublished Doctoral Dissertation, Department of Psychology, University of Edinburgh.

Wishart, J.G. & Bower, T.G.R., 1981, "A normative study of infant object concept development under three conditions of hiding," Manuscript submitted for publication.

Young, R.M., 1976, Seriation by Children: An Artificial Intelligence Analysis of a Piagetian Task, Basel, Birkhauser.

ANALOGY GENERATION IN SCIENTIFIC PROBLEM SOLVING

John Clement
Department of Physics and Astronomy
University of Massachusetts
May, 1981

A number of researchers have discussed the important role of analogical reasoning in science and education [1-12]. This paper describes research on the spontaneous use of analogies in problem solving by scientifically trained subjects. This occurs when the subject first spontaneously shifts his attention to a situation B which differs in some significant way from the original problem situation A, and then tries to apply findings from B to A. This is difficult for many people to do, possibly because it involves breaking out of the assumptions built up in considering the original problem. As a result, although spontaneous analogies are a more naturalistic phenomenon to study than provoked analogies, they are difficult to capture and record. However, by intentionally focusing on subjects who are known to have done creative work in the past, a number of such cases have been documented.

Ten experienced problem solvers were interviewed on a variety of problems. Most were video-taped. The subjects were advanced doctoral students and professors in technical fields. The findings summarized here are based on detailed protocol analyses of six of the problem solutions from this group that included the most significant uses of analogies. This brief paper concentrates on examples from the protocol of a single subject.

The first finding is that: spontaneous analogies have been observed to play a significant role in the solutions of a number of scientifically trained subjects. Solutions have lasted up to 90 minutes and some include reasoning patterns that are very complex. This complexity has led to a research focus of working toward a macro-level theory of the dynamic processes by which analogies are generated, evaluated, and applied. This is an appropriate initial strategy for mapping out a complex domain of processes about which little is known. From transcript analyses the general hypothesis was formulated that the following processes are fundamental in making an inference by analogy: [2]

(1) Given the initial conception A of an incompletely understood situation, the analogous conception B is generated, or "comes to mind";

(2) the analogy relation between A and B must be "confirmed";

(3) conception B must be well understood, or at least predictive; and

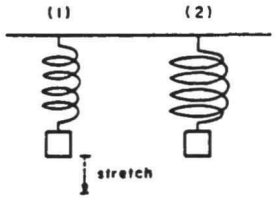
(4) the subject transfers conclusions or methods from B back to A.

This hypothesis is consistent with our observation that many successful solutions by analogy are not "instant solutions". Analogies are often proposed tentatively, and processes (2) and (3) especially can be quite time consuming. The last three processes can occur in any order, and subjects are often observed to move back and forth between them several times while gradually completing each step. This suggests that the subjects do not use a simple, well-ordered procedure for controlling their solution processes at this level. This paper focuses on steps (1) and (2). As will be shown there appear to be not one, but several ways of generating analogies, and several ways of confirming them.

EXAMPLE OF A SOLUTION CONTAINING ANALOGIES

Five subjects have generated analogies in thinking aloud about the following problem:

Fig. 1



Spring Coils Problem. A weight is hung on a spring. The original spring is replaced with a spring made of the same kind of wire, with the same number of coils, but with coils that are twice as wide in diameter. Will the spring stretch from its natural length, more, less, or the same amount under the same weight? (Assume the mass of the spring is negligible compared to the mass of the weight.) Why do you think so?

This problem was given to seven subjects. Four attempted to relate the problem to the analogy of a bending rod, as in the following verbatim, condensed transcript:

(1) S2: Um, I have one good idea to start with. It occurs to me that a spring is nothing but a rod wound up, uh, and therefore maybe I could answer the question for a rod. (Draws fig. 2)... I have a strong intuition, a physical imagistic intuition that this (rod a) will bend a lot more than that (rod b) will. In fact, the intuition is confirmed by taking it to the limiting case. It becomes intuitively obvious to me that as one moves the weight closer and closer to the fulcrum that the thing will not bend at all.

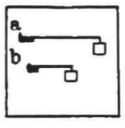


Fig. 2



Fig. 3

S2 goes on to infer that if the rod situation is truly analogous, the wider spring will stretch farther. Here the subject is able to achieve a high degree of certainty about the behavior of the rods (process 3 above). He reports doing this on the basis of what he calls physical intuition and by thinking about an extreme case, giving us reason to suspect that he is using some type of imagistic simulation process. But he is uncertain as to whether he can confirm the idea that the spring and the rod are analogous.

(2) S2: But it occurs to me that there's something clearly wrong with that metaphor because...its slope [the bending rod's] would steadily increase as you... went away from the point of attachment, whereas in a [stretched] spring, the slope of the spiral is constant... I don't see how that could make the bow go away; just to wind it [the rod] up. Damn it! [13]

He spends a large part of his 45 minute solution trying to resolve this issue. This transcript and others indicate that processes 1 through 4 above can indeed take place separately. S2 has apparently completed processes 1, 3, and 4 so far.

METHODS FOR CONFIRMING ANALOGY RELATIONS.

Determining a match between key relationships in both situations is the first and most obvious method for confirming analogy relations (process 2 above) [5]. Thus subject S2 above is worried because he cannot obtain a match between the changing slope in a bending rod and the constant slope in a stretched spring. However, other confirmation methods are also possible.

Confirmation via bridging analogies. Rather than throwing out the rod analogy, S2 proceeded to generate a second related analogy: the "zig-zag spring" shown in fig. 3. Such subjects are observed to generate an intermediate case when they refer to a situation that has aspects in common with two previous situations A and B. It is hypothesized that S2 attempts here to form a cognitive bridging analogy which links his conceptual frameworks for the rod situation and the original spring situation.

Figure 4 shows how such a bridging analogy can be effective [2]. The link labeled 1 represents the initial tentative analogy relation conjectured to exist between conceptions A and B. Here A is the poorly understood initial problem situation and B is a well understood situation. Inadequately vs. well-understood conceptions are represented by dotted vs. solid squares, respectively, and tentative vs. confirmed analogy relations are represented by dotted vs. solid links between squares, respectively. Figure 4 shows how the subject might establish a confirmed link between A and B by bridging back to conception A via conception C. If the analogy links (2) and (3) are confirmed (with respect to the same salient relationships between variables), then A can become well understood and become analogous to B, since under the above conditions, A being analogous to C and C being analogous to B means that A is analogous to B. We call this analogical transitivity. It should be emphasized that since "confirmed" generally means "intuitively compelling" rather than "proven" in this context, analogical transitivity is considered a form of plausible reasoning which does not lead to conclusions carrying the force of a logical implication. This diagramming system also allows one to construct macro-level "maps" of hypothesized cognitive processes occurring during complex solutions involving many analogies.[3]

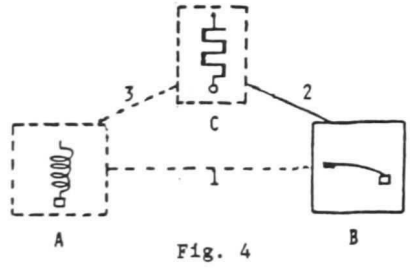


Fig. 4

A second bridge. Unfortunately, at this point the subject still could not reconcile the bending going on in the zig-zag spring with the lack of change in slope in the original helical spring (link 3 is unconfirmed), so his initial attempt at a bridge failed. However, he later generates a second, more successful attempt at a bridge in the form of an analogy to a polygonal spring. He is confident that a spring with hexagonal coils would not be essentially different from one with circular coils, and this leads him to a really new insight:

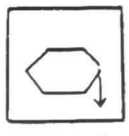


Fig. 5

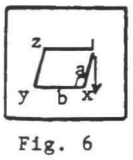


Fig. 6

(3) S2: Aha!... What if I start with a rod and bend it once (makes bending motion with hands) and...bend it again.. Clearly there can't be a hell of a lot of difference between the circle and say, a hexagon...(Draws fig. 5) Now that's interesting. Just looking at this it occurs to me that when force is applied here, you not only get a bend on this segment, but because there's a pivot here, you get a torsion effect...Aha! Maybe the behavior of the spring has something to do with twist forces... Let me accentuate the torsion force by making a square (draws fig. 6) where there's a right angle. Now...I have two forces introducing a stretch. I have the force that bends this...segment [a] and in addition I have a torsion force which twists at vertex, um, x...Now I feel I have a good model of a spring... Now making the sides longer certainly would make the [square] spring stretch more.

I: How can you tell?

S2: Physical intuition...and also recollection.. the longer the segment (moves hands apart) the more the bendability (moves hands as if bending a rod)... Now the same thing would happen to the torsion I think, because if I have a longer rod (moves hands apart), and I put a twist on it (moves hands as if twisting a rod), it seems to me--again physical intuition--that it will twist more... again, now I'm confirming that by using this method of limits. As I bring my hand up (moves right hand slowly toward left hand) closer and closer to the original place where I hold it, I realize very clearly that it will get harder and harder to twist... And my confidence is now 99% [that the wide spring stretches more]...I feel a lot better about it. [14]

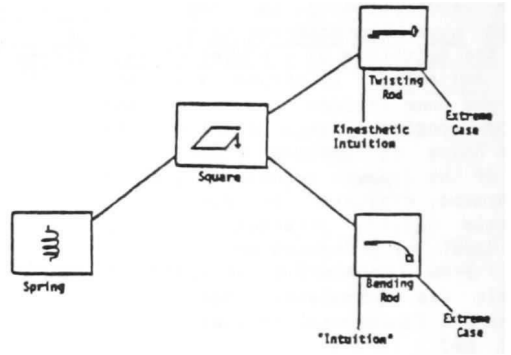


Fig. 7

Here he is able to firmly connect the original spring to the bending rod case via the bridging analogy of a polygonal spring. In addition, considering the polygonal spring triggers the recognition of a torsion effect. Thus, in the subject's final understanding of the spring, the spring is linked via the intermediate square spring case to two simpler cases, the twisted rod and the bent rod, as shown in fig.7. The torsion factor is an important insight, because not only is it true that wider springs stretch farther, but in fact the force provided by a helical spring is primarily due to torsion rather than to bending.

In summary, two major processes involved in confirming analogy relations have been identified: matching key features or relationships; and forming a bridging analogy.

ANALOGY GENERATION MECHANISMS

Analysis of transcripts has led us to propose the hypothesis that there are not one, but at least three types of analogy generation mechanisms: generation via an abstract principle, generative transformations, and associative leaps.

Generation from a principle. A plausible mechanism for generating analogies can be derived from the common situation in science where a single equation or abstract principle applies to two or more different contexts, such as a pendulum and an oscillating electrical circuit. This suggests that analogies may be formed by first recognizing that the original problem situation, A, is an example of an abstract equation or principle, P. The analogous situation, B, is then recalled or generated as a second example of principle P. However, although evidence for this pattern has been observed on occasion in interviews, little evidence of it was observed in the analogies generated for the spring problem. Instead, two other types of analogy formation processes appear to predominate, which I have called generative transformations and associative leaps.

Generative transformations. These occur when a subject modifies an aspect of problem A to create a new situation B. Examples of evidence for generative transformations from the present protocol are: (1) The subject refers to bending a rod into a polygon (protocol segment 3). (2) The subject referring to the spring as a "rod wound up" in the first line of the transcript indicates that the rod idea may have been generated by thinking about unwinding the spring. At another point he refers to the rod as an "unwound spring."

Associative leaps. In contrast to modifying the problem in a generative transformation, evidence for an associative leap occurs when the subject refers to an analogous situation B which is very different qualitatively in a number of ways from the original situation. The subject may also refer to "being reminded of" B. S2 generated evidence for several associative leaps in the middle of the protocol when he said: "I feel as though I'm reasoning in circles and I think I'll make a deliberate effort to break out of the circle somehow..., like rubber bands, molecules, polyesters...". apparently attempting to link the spring problem to other situations he knows something about. Although he was unable to use any of the associative leaps above effectively in this case, subjects have been observed to use this type of analogy generation technique successfully in other protocols. For example, one subject used an analogy to a U-tube to solve a problem about hydraulic forces in an apparatus whose shape and topology were quite different.

It is hypothesized that an associative leap takes place when an established conceptual framework for situation B in memory is activated by an association to some aspect of the original situation A, and that a generative transformation occurs when the subject focuses on an internal representation of the existing problem situation A and changes an aspect of it to create situation B. This leads to the prediction that an analogy generated via a transformation should more often be a novel invention (such as the hexagonal spring) and should more often contribute as a simpler case rather than as a more familiar case. Generative transformations and associative leaps have been the primary analogy generation methods observed by us so far [15].

METHODS FOR UNDERSTANDING A SITUATION AND FOR TRANSFERRING KNOWLEDGE FROM B BACK TO A

With regard to process 3 above, the requirement that conception B must be predictive or well understood, we note briefly that this can be achieved via factual knowledge, physical intuition, analysis in terms of a theory, or (recursively) via another analogous case C. Some methods for applying knowledge from B to A (process 4 above) are: (1) transferring a prediction directly from B to corresponding variable relationships in A; (2) transferring a partial understanding of certain variable relationships, which with further analysis can lead to a prediction in A; and (3) transferring a method of attack from B to A. [20]

EXTREME CASES AND PHYSICAL INTUITION

Minimizing or maximizing a feature of the problem sometimes makes the problem easier to analyze, and we call this using an extreme case. Extreme cases seem to be generated primarily via generative transformations or problem operators. Interestingly, the apparent function of many of the extreme cases observed so far has been to enhance the subject's use of physical intuition in the form of imagistic simulations. S2 indicates that his final understanding is based at the lowest level on such physical intuitions. This suggests that certain relationships between forces and other physical variables such as "bending" can be represented at a deep level in terms of imagistic intuitions rather than abstract principles or equations. (See ref. [3]).

CONCLUSION

Further research is needed in order to evaluate and add to the results of this exploratory study. A number of basic concepts for analyzing patterns of analogical reasoning have been proposed, including: the generation of analogies via transformations and associative leaps; the evaluation of analogy relations via the formation of bridges and the matching of key relationships; and the understanding of situations via the use of extreme cases which can enhance physical intuitions. Recursive combinations of these processes can account for many of the patterns observed in other complex solutions involving a number of linked analogies. Many solutions by analogy are not "instant solutions", but a more extended process of conjecture and testing. This gives us reason to believe that some of these processes are learnable, rather than being exclusively a product of "genius", and that developing students' abilities to use generative transformations, leaps, and bridges may be possible and desirable.

When a transformation leads to a confirmable analogy, we call it a conserving transformation since it conserves the salient relationships in the problem. In a broader sense, conserving transformations appear to play a fundamental role at different levels in physics, mathematics, technological invention, and music [16-19]. Conserving transformations appear to be an important cognitive process worthy of further investigation.

In the case of S2, the bending rod analogy served as a first order model which gave him an initial handhold on the problem. Persistent criticisms and transformations of this model during his vigorous 45 minute solution eventually led him to evolve a much better model in the form of a square spring with torsion effects. Thus, sophisticated uses of analogy in relatively difficult problems can involve a repeated conjecture, criticism, and modification process that can produce chains of successively more powerful analogies. Analogous cases can either play a temporary heuristic role in helping to generate conjectures during the solution, or they can play the more permanent role of a model in the final solution, or both. Certain parallels between these processes and processes of science described

in [6-8,16], among others, suggest that further research along these lines may be of interest to those studying the processes of hypothesis formation and model construction in science.

NOTES

[1] Brown, J.S., Collins, A., and Harris, G., "Artificial Intelligence and Learning Strategies". in H.F. O'Neil (ed.), Learning Strategies. New York: Academic Press, 1978.

[2] Clement, J., "The Role of Analogy in Scientific Thinking: Examples from a Problem Solving Interview", technical report, Department of Physics and Astronomy, University of Massachusetts at Amherst, 1977.

[3] Clement, J., "Spontaneous Analogies in Scientific Problem Solving", working paper, Department of Physics and Astronomy, University of Massachusetts at Amherst, 1981.

[4] Dreistadt, R., "An Analysis of the Use of Analogies and Metaphors in Science". The Journal of Psychology, Vol 68, 1968.

[5] Gentner, D., "The Structure of Analogical Models in Science", technical report, Bolt, Beranek and Newman, Inc., Cambridge, MA, 1980.

[6] Hesse, M., Models and Analogies in Science. University of Notre Dame Press, Notre Dame, 1966.

[7] Kuhn, T.S., "Second Thoughts on Paradigms", in The Essential Tension. University of Chicago Press, Chicago, 1977.

[8] Lakatos, I., Proofs and Refutations: The Logic of Mathematical Discovery. Cambridge University Press, Cambridge, England, 1976.

[9] Rissland-Michener, E., "Understanding Understanding Mathematics". Cognitive Science, Vol 2, 1978.

[10] Polya, G., Mathematics and Plausible Reasoning. (2 Vols) Princeton University Press, Princeton, New Jersey, 1954.

[11] Rumelhart, D. and Norman, D., "Analogical Processes in Learning", report no. 8005, Center for Human Information Processing, University of California at San Diego, 1980.

[12] VanLehn, K. and Brown, J.S., "Planning Nets: a representation for formalizing analogies and semantic models of procedural skills", to appear in R.E. Snow, P.A. Frederico, and W.E. Montague (Eds), Aptitude Learning and Instruction: Cognitive Process Analyses. Lawrence Erlbaum Associates, Hillsdale, NJ, 1979.

[13] His meaning can be clarified here by noting that a bug would experience a slope of constant steepness in walking down a freely supported, stretched spring and a slope of increasing steepness on a bending rod. He later says that by "spiral" here he meant "helix".

[14] Two of the subjects in the sample, including S2, had heard that the author was interested in studying analogies prior to their interview. However, the subjects did not know which problems in the interview were amenable to solutions by analogy. After the interview, S2 was asked if he thought this prior knowledge had affected this solution and he felt strongly that he had simply solved the problem in the way he normally would. Most importantly, none of the subjects had communicated with the author concerning any of the theories or concepts developed in this paper, such as transformations, leaps, bridges, etc.

[15] Bridging analogies are generated under the multiple constraints of being similar to two situations. This leads us to predict that bridging analogies should ordinarily be generated via a transformation rather than a leap, and this is so far consistent with our data. In future reports we will show how non-generative transformations can also be involved in confirming an analogy relation.

[16] Cohen, I.B., The Newtonian Revolution: With Illustrations of the Transformation of Scientific Ideas. Cambridge U. Press, 1980.

[17] Krueger, T., Imagery Pattern in Creative Problem Solving. Encina Press, Las Cruces, New Mexico, 1981.

[18] Shepard, R., "Psychophysical Complementarity", in M. Kubovy and J.R. Pomerantz (Eds) Perceptual Organization, Lawrence Erlbaum Associates, Hillsdale, NJ, 1979.

[19] Rissland, E.L., "Example Generation", COINS Technical Report 80-14, University of Massachusetts at Amherst, 1980.

[20] The four criteria for a useful analogy on the first page apply to the strongest use of an analogy where a prediction is transferred. Transfer of a partial understanding or possible method of attack can occur without criteria 2 and 3 being fully satisfied.

Understanding Design

Gerhard Fischer and Heinz-Dieter Boecker
Research Group on Man-Machine Communication
Computer Science Department, University of Stuttgart

Abstract

Design is concerned with how things ("artifacts") ought to be in order to attain goals and function. To gain a deeper understanding of the nature of design, we have studied design processes in several different domains: software engineering, creative writing, building models with technical kits and constructing learning environments.

1. The nature of the design process

Designers try to shape the components of new structures. In general the designer is facing an ill-structured problem in a semantically rich domain. An emerging design is incorporated in a set of internal and external memory structures (eg on paper, in computer memory). The problem solving process for this task can best be characterized as a dialectical information gathering process switching back and forth between **creating** and **understanding** subprocesses. At each stage in the design process, the partial design (reflected in the descriptive structures) serves as a major stepping stone and as a stimulus for suggesting what should be done next (see ALEXANDER 1964 and SIMON 1969).

There is no doubt that there is domain specific design knowledge: an architect knows a lot about building materials and how to combine them, a software engineer knows many basic algorithms and how to use them in constructing software components. The more interesting question in research about design is whether there is also some general design knowledge, which is applicable in many different domains.

2. Universal concepts for design processes

Crucial processes in **creating a design** are:

1) to develop a language for specifying goals and a set of intermediate concepts (eg in the case of technical construction systems: a steering unit, a differential)
2) to deal with sets of possible worlds (ie we have to consider design alternatives)
3) to incorporate the emerging design in a set of external memory structures
4) to record the relevant parts of the design path (and not only the design product)
5) to create low-cost modifiable models which help us to replace anticipation by analysis and experimentation

Understanding (analysis) of existing designs is relevant for the following reasons (especially if we regard the design primarily as an error correcting process):

1) it is a prerequisite for modification (which is necessary because it is not possible for the designer nor for the potential user to foresee all of the opportunities for which a complex system is used).
2) the insufficiencies of existing prototype models have to be uncovered; the surface manifestations of problems have to be traced back to their causes.
3) contemplating and experimenting with what has been done so far can stimulate new ideas, new generalisations, new simplifications and new uses for developed parts. New possibilities may become visible through perception of unplanned relationships.

Computer science has developed a set of well known (eg top-down, bottom-up, stepwise refinement) and less known (eg use of stepping stones, progressive constraints, structured growth, linguistic uniformity, uniformity of interaction, metadescription, context-sensitive descriptions, filters) concepts of design.

3. Technical construction systems and systems for man-machine communication

To gain a better understanding for an epistemology of design, we have studied design processes in several domains: software engineering (BOECKER, FISCHER and GUNZENHAEUSER 1980), creative writing (FISCHER 1980), technical construction systems (FISCHER and BOECKER 1981) and learning environments (FISCHER, BROWN and BURTON 1978). We have verified and extended the basic set of thought processes underlying design and the organization of those processes which serve as guidelines for the implementation of design support systems.

In our own work we have chosen a specific technical construction system, namely FISCHERTECHNIK (FT). FT can be enjoyed by "children of all ages". It allows the designer to construct starting with a universal building block - very simple models (eg a car, a crane etc) as well as completely functional models (eg a technically exact model of an assembly line for car manufacturers). FT is not only an excellent toy (it gained the "Oscar de Jouet" in 1972 for being elected the best toy of the year) but a realistic design tool which allows to build artifacts of comparable complexity relative to big software systems.

FT is a good example of a design tool that has the property **"no treshold, no ceiling":** easy things are easy to do and difficult things are possible to do. The most important feature of FT is probably the mechanism to join individual building blocks (which is uniform and universal and can be undone in all cases). It provides the basis for an environment which allow us to study sophisticated design processes with great similarity to software engineering.

The other major research area in our work for an epistemology of design has been the **design of man-machine communication systems.** We will make use of the substantial computational resources at the hardware level (which will become available in the near future) to create a symbiotic environment for man and machine which combines the advantages of both to cooperatively solve design tasks which neither of them could solve alone. The main difficulty in this domain is our lack of a theory which is predictive enough to provide a complete set of design criteria. We try to make use of the knowledge about the human information processing system (eg limited STM, good visual chunking methods) to come up with some principles (eg consistency and adequate default assumptions) for the design of these systems. The few derivations we can make from these principles, however, immediately lead to serious design conflicts. To point out a specific example: if we want to delete a file (out of a large set of successive versions), the default option should be to delete the oldest version whereas if we print or edit a file the default should be the newest version. At the surface level this is obviously an inconsistency. Only if the user has an explicit conceptual model of these processes (eg about the semantics of "delete", "print" and "edit") this seems to be the obvious way.

One of the long range goals of our research is to construct computer-based **convivial tools** (ILLICH 1973). These are tools that give the user as much control as possible so that he can do what ever **he** wants to do. We consider Fischertechnik a good example of a convivial tool (eg compared with a plastic car which can never be modified).

Our **empirical research** verified some of our intuitions, namely

1) there exist common features for design processes in different domains
2) design is an incremental activity; a partially completed product is important and useful to gain a deeper understanding of the problem;
3) the major difficulty in design tasks is not to understand the intrinsic semantics of individual building blocks but to predict how an assemblage of such components will behave
4) it is important for the understanding of an artifact that we have access not only to the final product but to the important parts of the evolutionary path which led to it
5) the language developed in Computer Science and specifically in AI is adequate for the description of design processes in different domains.

4. A computer system to support design processes

Many design problems (including social planning, VLSI design, software engineering) have demonstrated to us that there are inherent limitations to the complexity that the unaided designer can control in any situation. Our system INFORM can best be described as an integrated, knowledge-based design environment (primarily used for software creation, maintenance and modification) using multiple windows to provide uniform, direct interactive facilities to the user. We expect that systems of this sort (see also GOLDSTEIN and BOBROW, 1980) will make major contributions to cope with the problems of constructing large software products by keeping the whole programming process in the machine (and not only the result) and by allowing to extend our specification techniques with prototyping.

5. Conclusions

Our goals doing this research were: to gain a better understanding of design (a cognitive science task) and to use this understanding to build systems which suport the design process (a cognitive engineering task). We believe that domain independent design knowledge and skills exist and that design concepts are the right vocabulary to describe complicated processes in problem solving, expert systems research and software engineering.

References

Alexander, C. (1964): "The synthesis of form", Harvard University Press

Boecker, H.-D., G. Fischer and R. Gunzenhaeuser (1980): "Die Funktion von integrierten Informationsmanipulationssystemen in der Mensch-Maschine Kommunikation: Projekt INFORM", MMK Memo, Institut fuer Informatik, Universitaet Stuttgart

Fischer, G. (1980): "Integrated Information Manipulation Systems -- A cognitive view", Proceedings of Coling 1980, the 8th International Conference on Computational Linguistics, Tokyo

Fischer, G. and H.-D. Boecker: "The nature of design processes and how computer systems can support them", submitted to IJCAI 81

Fischer, G., J.S. Brown and R. Burton (1978): "Aspects of a theory of simplification, debugging and coaching", in Proceedings of the 2nd National Conference of the Canadian Society for Computational Studies of Intelligence, Toronto, pp 139-145

Goldstein, I. and D. G. Bobrow (1980): "Descriptions for a programming environment", Proceedings of AAAI, pp187-189, Stanford

Illich, I. (1973): "Tools for Conviviality", Perennial Library, Harper and Row, New York

Simon, H. (1969): "The sciences of the artificial", especially chapter 3: "The Science of Design", MIT Press, Cambridge, MA

System Properties of American Law: Constraints on Applying Cognitive Science

Janet L. Lachman and Roy Lachman

University of Houston

One measure of the value of a science is its impact on other sectors of society. Cognitive theory has undergone unprecedented growth in the past decade and there are many areas of potential application. One of these is the institution of law. However, law is an intricately structured system, and efforts to apply cognitive science research findings in legal settings will come face to face with a fundamental reality: the systems properties of the law, and not the extent of our expertise, will determine the nature and extent of such contributions. Law is an intellectual enterprise, and much cognitive data appears strikingly pertinent to its operation. However, substantial differences between scientific and legal outlook can result in different judgments of pertinence. Many of our colleagues, even those who have frequently testified as experts, have been puzzled and even outraged by the exclusion of some of their most salient inputs. One who understands the system is likelier to produce information that will be welcomed, or at least appreciate the factors leading to rejection when it occurs. This paper will consider certain of the objectives of a trial that determine what expert testimony is likely to be admitted, and to suggest four points at which cognitive expertise might find entree to institutions of law.

There is much more to law than the trial of cases, but the trial is a central feature of our legal system. It is our ultimate tool for resolving disputes and for imposing criminal sanction. It is also the most common setting for the introduction of non-legal expertise. The nature and extent of expertise that is admitted is determined by the purposes of the trial and the values the system has placed on certain competing purposes. A trial is essentially a determination whether the facts of a particular case fit a given pattern, where the pattern has legal consequences. For example, if a physician commits malpractice and thereby injures his patient, the legal consequence is that he owes the patient money. "Malpractice" is essentially a legal pattern consisting of failure to provide the level of care the doctor implicitly promised when he undertook to treat the patient. When a patient sues his doctor for malpractice, the trial will be an effort to find out what level of care the doctor implicitly promised, and whether the treatment fell below that level; whether the patient was really injured, and if so, whether the doctor's conduct caused the injury. These facts are established on the basis of the evidence admitted at the trial, which includes the testimony of witnesses, expert and otherwise. In a jury trial, the jury determines the facts, within certain limits, and applies the law as instructed by the judge. In a bench trial, the judge finds the facts and applies the law. In either case, the factual determinations are made in a context of competing values, of which achievement of truth is only one.

The legal standard of truth is not, and never has been, absolute. For centuries, courts have decided the truth of claims where absolute accuracy is unobtainable. What is obtainable in our adversary system is procedural fairness; and as our system has evolved, procedural fairness has become the measure of the truth achieved. One dimension of the fairness of the system is the speed with which a result is reached: "justice delayed is justice denied." Moreover, the perception of fairness is a significant dimension of the system; if people do not trust institutional methods of dispute resolution, they will turn to non-institutional remedies. And of course, a major purpose of a justice system is to avoid resort to such remedies.

The presiding judge at a trial is attempting to structure a procedure that achieves the best compromise of these sometimes competing values. He must not prejudice the interests of a party; he must not deprive a party of a fair trial. But the system does not require him to give any party the best trial possible. There are many different judicial decisions that are consistent with fairness; and typically the judge will not be reversed as long as his decisions have not denied a party a fair hearing. Among the decisions the judge makes is whether any witness, expert or otherwise, shall be heard. He must admit testimony only if it would be unfair to exclude it. He will admit testimony if he thinks it will help the jury decide the issue it must decide, sufficient to offset the costs in court time.

The trial, then, is a method for determining disputed facts by means of structured procedures. The distinction between "facts" and "law" is a fundamental one. "Facts" in the malpractice lawsuit previously mentioned are the nature of the doctor's implied promise, and whether his conduct caused the patient's injury, etc. The "law" encompasses virtually all else, including the trial procedures themselves. Witnesses are supposed to testify to facts only; expert witnesses are permitted to express opinions on factual matters. However, no witness is supposed to testify to matters of law. Consequently, testimony that is relevant to an ultimate fact in issue tends to be freely admitted; but testimony that has implications for the fairness of the procedures is less welcome. Certain types of psychological expertise, particularly, tend to relate rather intimately to the conduct of the trial itself, and in these areas courts may be surprisingly resistant to hearing what the expert has to say. We shall return to this point.

How, then, does cognitive expertise exert an impact on the legal system? The mechanics of impact are multifarious; but there appear to be four "points of system operation" at which cognitive research might be received. The first of these might be called "procedural reform". Procedural reform involves changes in procedural rules themselves. This is the level at which research on the effects of jury size impacts the system, for example. Procedural reforms are basically a legislative prerogative, and research concerning procedural reforms is most likely to impact the system in legislative committee hearings, direct lobbying, and the like. If cognitive research on judicial decision-making has procedural implications, this would be the appropriate level for its introduction.

The second point at which cognitive science is likely to impact the law is at the traditional level of assisting in fact finding. This form of impact is achieved by the familiar means of testifying at trial, regarding some area of expertise that has become an issue in a particular lawsuit. One example is testimony about the effect of fatigue on reaction time in an automobile accident case.

The third point is subtle, falling somewhere between the first and second. We shall label it "judicial control processes", but the label is not self-explanatory. At this level, the cognitive scientist appears to be testifying to facts, in the conventional manner, but the testimony has implications for trial procedures. As previously mentioned, testimony of this type may be resisted.

Cognitive research specifically designed to have legal significance may impact the legal system at this level, but it requires a sympathetic judge for it to be heard. One example is eyewitness identification research. The cognitive scientist is called, usually by the defense in a criminal case, to tell the jury the results of research on eyewitness identification. The researcher's testimony is received as "fact" testimony, nominally to help the jury evaluate the credibility of the eyewitness. However, it has wider implications, because it suggests that juries need assistance in evaluating the general credibility of eyewitnesses -- not just the eyewitness in a particular case. Juries are presumed to possess without advice proper perspicacity on such matters. The judicial impulse may be to exclude the testimony, perhaps on the basis that it will not help the jury. And in one sense, it won't: if the prosecution's case depends heavily on the eyewitness, how can it help the jury to be told that their best source of information is unreliable? No comparable problem exists where the testimony is more traditional; it need not unsettle the judge if a witness swears that his research shows that tired people react slowly. Despite pressures to exclude expert testimony on the reliability of eyewitnesses, there is a trend toward admitting it. This trend may reflect a factor that countervails judicial resistance: only the defendant can appeal in a criminal case, and it is usually defendants who seek to introduce expert testimony to counter the effects of an eyewitness. By giving the defendant the benefit of every doubt, the judge minimizes the occasions on which his decision is reversed on appeal.

A second example of cognitive science research impacting the legal system at this third level is work on hypnosis. Many witnesses have proved able to recall additional and important detail under hypnosis that they could not recall otherwise. There are important legal issues involved in the use of hypnosis: Can the witness meaningfully swear the oath to tell the truth while hypnotized? Can the witness swear before being hypnotized, and comply with the oath afterward? Can the witness be meaningfully cross-examined? Cognitive research on hypnotism may prove considerably more valuable to our legal system than that on eyewitness identification, for it potentially adds to the fact-finding arsenal. It has also met with initial resistance; like eyewitness identification reliability, it has implications for the conduct of the trial itself. However, if the nature of hypnosis can be understood, the legal issues can be addressed rationally and the system eventually will encompass it.

The final area of potential impact might be called "indirect inputs". Empirical data available in cognitive science that are rejected at the other levels can nevertheless be used informally by attorneys. A well-known example of impact at this level comes from social psychology, where empirically-based theories of small-group interaction have been used to assist lawyers in the jury-selection process. Comparably, a cognitive scientist whose formal testimony on eyewitness identification has been rejected by the trial judge may nevertheless suggest to the attorney the most vulnerable aspects of the eyewitness' account, and these suggestions may be used to advantage in cross-examination.

By far the smoothest entry to the legal system is in the conventional role of expert on a disputed issue of pure fact. Much of our research has implications for important areas of the law, which would be quickly noticed by the legal profession if it were properly packaged. For example, psycholinguistic research on language comprehension is typically done using narrative prose. However, reading the instructions on equipment, medicine bottles and the like also involves language comprehension -- and also figures in an important class of products liability cases. There may be some questions that could be meaningfully addressed in either context; if so, why not use the one with practical utility? Incidental learning and recognition memory, two familiar areas of memory research, are significant factors in trademark infringement cases. Some of these research questions might be addressed by using real or simulated trademarks as stimuli; and the researcher would become a potential contributor to the patent and copyright bar. Research on cognitive development could suggest when a child is old enough to make informed decisions on matters affecting his own welfare, such as medical treatment (including abortion), which parent shall have custody, etc. Psychological and psycholinguistic inquiry into the concept of intentionality potentially has profound implications. Many additional areas of research could be suggested that would be both quality cognitive science and oriented toward significant issues of law.

We have suggested four distinct ways that cognitive scientists can seek to make contributions to our legal institutions. The reception will be different depending upon the point of impact; the cognitive scientist who understands the properties of the legal system will be in a better position to comprehend its response to his proffered contribution.

Writing with a Computer

Ira Goldstein
Xerox Palo Alto Research Center
Palo Alto, Cal. 94304

Abstract: This essay conjectures that an author's planning process will be facilitated by a tool that represents his plan at various levels of abstraction as a network of subgoals, with the subgoals not necessarily restricted to a linear order. Machine reasoning on such structures has been explored in artificial intelligence research: our proposal is to make these structures available to the writer as a calculus for representing his essays and to use the computer as an interactive editing tool to manipulate them.

Writing and Planning

Machine planning: If we examine the literature on machine planning [1], it is apparent that data structures for plans must have a capacity to represent plans at various levels of abstraction. Otherwise the planning program is prematurely mired in unnecesary detail. Furthermore, at any particular level, the representation should minimize constraints on the order in which goals must be accomplished: it should be possible to express that two goals must be satisfied at a certain point in the plan, but in an unspecified order. Again, this is to avoid having the planning process become prematurely committed to a particular solution. It should also be possible to express at a more concrete level that the goals in question must be accomplished in a fixed order to satisfy constraints that arise at that level.

Problems of writing: These characteristics of the planning process apply to organizing an essay just as much as they apply to a robot choosing a path to some destination. An author has a variety of topics to present--his goals--and ought not to be prematurely tied down to a particular order early in his planning process. Furthermore, he should be able to consider the organization of these topics independently of their details.

Limitations of paper: Authors can meet the demands of planning with pencil and paper. They may work with outlines to deal with their material abstractly and 3x5 cards to avoid premature commitment to a particular ordering. However, as a medium for planning, paper has a number of difficulties. Exploring alternatives is cumbersome. It is difficult to maintain two versions of an outline or a file of cards. Backing up to a previous version is equally difficult, especially if some changes have been made that apply to both the new and old versions. Avoiding premature commitment to a particular organization is hindered by the linear nature of prose. Constraints regarding length, figures or citations must be remembered by the author with no help from the medium itself.

Virtues of the computer: The computer is a more flexible medium than paper for planning because it supports data structures that capture the nature of plans better than linear strings of text or files of index cards. The computer can represent alternative versions with shared structure, maintain a history of previous versions, represent nonlinear organizations of goals by means of networks, and express constraints as programs that monitor the evolving plan. These data structures can, of course, be sketched on paper, but they rapidly become too complex to edit easily. The computer can serve as an editing device that simplifies these data structures by means of filtered views and presents them graphically to an author so as to make them comprehensible and easy to edit.

An Example

To exemplify this, I will use my own experience in writing the introduction to this essay. While it is short, its generation nevertheless required the solution of a variety of typical writing problems. I shall show how the computer was employed as a tool for coping with these problems.

> This example is offered with an awareness that the value of the computer as a planning aid increases in proportion to the complexity of the writing task. In this respect, planning a journal article or a book would be a more compelling example than planning an introduction to a short essay. However, such an example would also be more complex and time-consuming to present. Hence, I have chosen a simple, but real case to present. The reader is asked to generalize this example to writing problems that he has encountered, especially in the context of longer and more complex documents.

As with any introduction to a research article, my subgoals were to present a brief statement of the problem that that I was attempting to solve, the nature of previous solutions, their limitations, the particular solution that I was proposing, and the evidence for this solution. The problem that I faced was how much to say about each of these topics and in what order.

Had I pursued this task with pencil and paper, I would typically have written several drafts of the introduction. The drafts would have included changes to both organization and the content of individual paragraphs. I might also have created and revised an outline of topics to be discussed, sometimes to serve as an initial plan, sometimes to analyze an existing draft.

Instead, I wrote this introduction using a writing environment implemented in PIE, a prototype personal information environment for the representation of designs [2]. Figure 1 is a graphic representation of the top level network that I constructed to represent an early plan for this introduction. The node labelled *Introduction* represents the main goal. It is preceded by the *Abstract* node and followed by the node that represents the goal for this section. The box in boldface linked to the *Introduction* node is the plan for accomplishing this goal. It consists of four subgoals that must precede the statement of my particular solution, but are as yet unordered.

The first return that I obtained from using PIE is reflected in its ability to express and manipulate a nonlinear sequence of topics. Plan 1 did not commit me to a particular order for discussing *G4* through *G7*.

Figures 2 and 3 show two alternative refinements of Plan 1. These refinements differ in the order in which they propose to discuss the subgoals. They are similar in that *G6* has been eliminated in both. The basis for this decision is that it is not a topic of sufficient interest to the intended audience--the members of the cognitive science society. The details of *G6* still remain in the computer database and available for other discussions of this research.

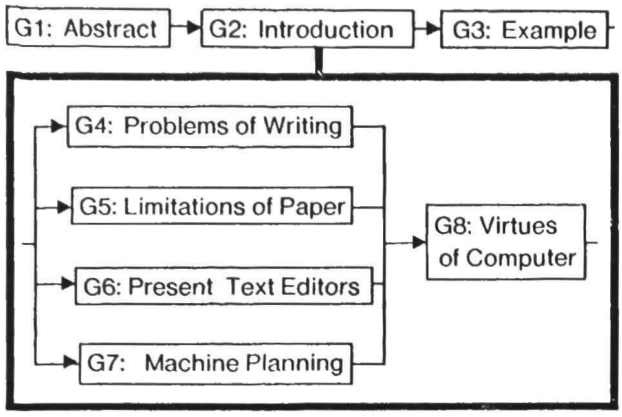


Fig. 1. Plan 1, an early plan in which the ordering of subgoals has not, as yet, been entirely determined.

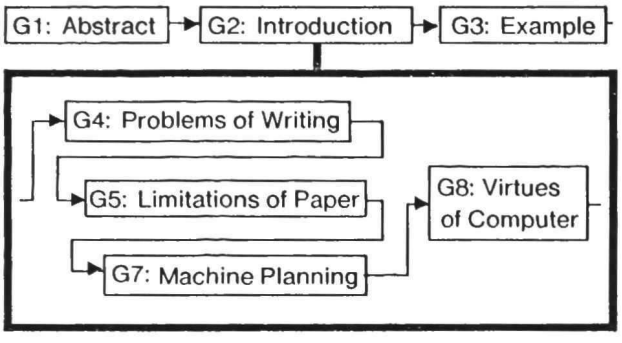


Fig. 2. Plan 2, a refinement to plan 1 in which G6 has been suppressed and the remaining goals ordered.

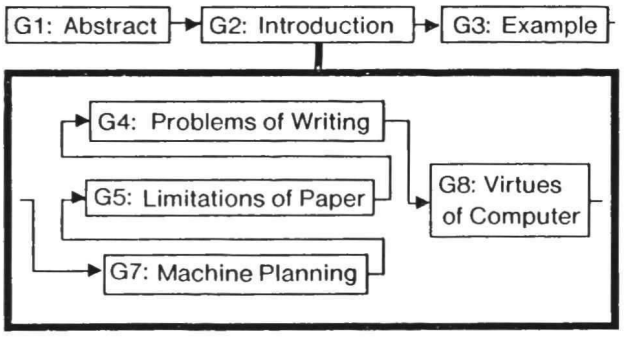


Fig. 3. Plan 3, an alternative to plan 2 in which a different order for the subgoals is chosen.

The ability to manipulate alternatives is the second return that I obtained from using PIE. With paper and pencil, multiple drafts present the difficulty that one cannot edit some paragraph common to two drafts and have the changes show up in both. Nor can one easily place two drafts side by side, with their differences highlighted. Both capabilities are present in a computer-based planning system. The link from *G4* in both plans can point to the same subnetwork; hence, changes to that network will be reflected in both plans. Analysis programs can examine two network structures and highlight their differences.

Ultimately, I chose Plan 3 based on a belief that the initial discussion of writing difficulties called for in Plan 2 was unnecessary because they were well known. The relevant difficulties of writing are mentioned three paragraphs later in the context of describing the application of machine planning to the writing task. The appeal of Plan 3 is that the discussion occupies less space, a significant virtue given a limit of 2000 words.

This leads to a discussion of a third virtue of PIE for document planning: the ability to assert constraints on the plan and have the computer monitor these constraints. PIE allows the designer--in this case, author--to assert predicates regarding properties of the network that he wishes to be monitored. One such predicate assigned to the node representing the document as a whole is that its entire length not exceed 2000 words. Another kind of predicate assigned to paragraphs that reference bibliographic material is that these citations appear in the bibliography. Similar consistency-related predicates apply to section and figure references. These predicates serve as reminders to the author and as a mechanism to preserve consistency among the various parts of the document being edited.

A fourth virtue of the system that is closely related to constraints is the ability to view the network through a filter. Here a constraint is imposed for the purpose of limiting the portion of the network that appears in a given view. Figures 2 and 3 exemplify one such filter: nodes that are not linked to a plan are suppressed. The result is that the node labelled *Present Text Editors*, which appears in Figure 1, is not present in these views. In general, a filter is a predicate on the network and results in some set of nodes and/or links being suppressed.

Objections

The reader has no doubt thought of objections to the claim that this approach will facilitate good writing. Here are three common criticisms, and possible replies.

Structures appropriate for representing machine plans are not appropriate for representing human plans.

Clearly some data structures appropriate for computers are entirely inappropriate as a calculus for people to employ for similar tasks. And any machine data structure is inappropriate if presented to people at too low a level of implementation. However, this is not to say that all formalizations of cognitive processes developed for machine problem solvers are useless for people.

I believe that the formalization of planning described here is useful because it is largely based on the nature of planning, rather than on the idiosyncracies of computers. As evidence for this, planning networks have been developed in the context of PERT charts to analyze and guide the planning of complex projects like the construction of ships. The similarities between these networks and the AI structures is far greater than their differences.

Conscious planning to the degree proposed here will act as a barrier to creative writing.

This is a version of the *centipede* argument--namely, that if a centipede thought about or planned his perambulation too carefully, he would fall over into a ditch in utter confusion. Better to just engage in the process without conscious examination.

Teachers of writing courses would take issue with the *centipede* argument. In such courses, students are taught many strategies for organizing their material, and the claim is made that attention to organizational issues repays the writer many times over at later stages of the writing process. Our position is that if planning is useful, surely the computer can provide a better mechanism than 3x5 cards.

We do not argue that a document should always be approached in a top-down mode. At certain times, the best strategy is to write a particular section in some detail before completing the plan for the entire document. Using the computer does not prevent this. The author can move from one level of the data structure to another at his choice. Figure 2, for example, might have been the first plan created by an author. Later, in searching for an alternative organization, he might step back and express the less committed plan of Figure 1, then refine it to the plan of Figure 3.

The overhead in using the computer is too high; better pencil and paper because it is easy to use and does not itself obstruct the writing process.

Unless careful attention is paid to the human factors of designing a planning editor, this objection is a telling one. Powerful planning structures are useless unless they are easy to manipulate and comprehend. The graphic display of the planning network used in the figures of this article constitutes an implicit proposal for a presentation mode that is easy to understand. The PIE system presently uses a non-graphic display of the network: one that requires more tutelage than the network diagrams, but is easier to implement. Research into the mental models that users have of such networks and into user-interface design that they find comfortable to employ is critical to the success of such tools.

However, even with a good interface, some planning overhead will remain. Hence, another reply to this objection is that only some documents justify the overhead. One-page memos do not. But our hypothesis is that for more complex documents, the planning overhead required by the computer can be less than the overhead required by pencil and paper.

A third response is that while authors are presently more familiar with pencil and paper than with computers, this will not be so a decade hence. For a large number of reasons, it is reasonable to predict that computers will become a universal technology, and that computer literacy will be as common a subject as penmanship. Hence, that part of the overhead due to unfamiliarity with computers is on the wane.

Extensions

We did not attempt to formalize the kinds of plans that an author might employ, beyond providing a means to represent subgoals and successor relationships. However, books on rhetoric and debate contain lists of such plans. This suggests that one might be able to define a set of generic planning schemata to represent different arguments such as argument-by-induction, argument-by-authority, and argument-by-deductive-proof. A given schema would contain slots for the various positions that a given argument requires: the axioms and logic of a proof; the set of examples for an inductive argument, etc. A writer could then expand a plan for a document as a series of instantiations of different schemata. Whether this would result in more coherent or compelling prose remains to be seen, but it would at least be useful as a device to articulate a formal theory of argument structure.

An application of the computer complementary to its use as a writing tool is its use as a reading instrument. The planning structures created by the author can serve as a roadmap and the computer can act as a device for examining these structures. Potentially, this increases the reader's ability to browse through material in idiosyncratic ways, rather than being tied to a fixed order of presentation selected by the author. It may also simplify the writer's task by eliminating the need to find a single linearization appropriate for all audiences. How comfortable a reader will be with nonlinear information structures and whether there are writing and reading skills uniquely appropriate to them are research questions that must be addressed.

Conclusions

This essay has proposed that planning is one writing skill that can be facilitated by the use of the computer as an design tool. Experiments to verify this proposal and research to investigate what other skills of writing might be facilitated by this medium remain for the future: what is certain now is that if we program these machines to imitate paper, as is the case for the current generation of text editing systems, we will never know if qualitative improvements in the handling of words can be obtained.

Bibliography

[1] Sacerdoti, E. *A Structure for Plans and Behavior*, Elsevier Computer Science Library, New York, 1977.

[2] Goldstein, I and Bobrow, D. "A Layered Approach to Software Design" to appear in Sandewall, E., Shrobe, H. and Barstow, D., *Interactive Programming Environments*, McGraw Hill, 1981.

CREATING PLEASANT PROGRAMMING ENVIRONMENTS FOR COGNITIVE SCIENCE STUDENTS

M. Eisenstadt, J.H. Laubsch, & J.H. Kahney
Open University, England

1. Introduction and background

This paper describes our current efforts towards the systematic improvement of a LOGO-like software environment called SOLO (Eisenstadt, 1978), which has been used by over 1500 Cognitive Psychology students at the Open University. SOLO is geared towards the manipulation of assertional data bases, and provides the students with a handful of easy to use primitives with which they can address some elementary problems of knowledge representation. Students login to one of our DECsystem-20s from a regional study centre, and experience about 10 hours of hands-on activity early in the academic year. Later on, they attend a residential summer school at which they can acquire an additional 30 hours of hands-on experience.

Although SOLO is a toy language in some respects, the total user environment has many features which make it ideal for providing the vast majority of our students (80% of whom are computer-naive) with their first exposure to computing. Among these features are a spelling corrector, syntax-directed editing aids, automataic display of data base changes as they occur, "undo" facilities, and an easily modifiable user-profile.

An in-depth analysis of our students' errors (Lewis, 1980) has led to an improved design to help ensure that beginners can write syntactically correct programs with a minimum of fuss. A microcomputer implementation, which uses screen-oriented syntax-directed editing (cf. Teitelbaum & Reps, 1980) is being piloted just prior to this conference.

Even with SOLO's extensive user aids and carefully pre-tested curriculum materials, our students still experience problems in writing programs which perform precisely as intended. Because of this, we have undertaken a detailed analysis of their programming behaviour. Our empirical studies, described in section 2, have highlighted the use of a small number of programming schemas by a large proportion of our students. These schemas serve as the basis for an automated debugging assistant, which is described in section 3.

2. The behaviour of SOLO programmers

As part of their SOLO activities at the beginning of the year, our students are asked to "write a program which makes the following inference: If someone is found to be guilty, then whoever that person works for is also guilty." In solving the problem, students are invited to invent their own data structures and algorithms. We analyzed a sample of 160 student programs to see if some underlying order could be found among a potentially large variety of databases and program structures. As it turns out, the programs written by these students are built from a handful of basic program schemas. These schemas are langauge-independent programming constructs such as FILTER, CONJUNCTION, SIDE-EFFECT, and GENERATE-NEXT-OBJECT, which are closely related to those found in the LISP "plan library" of Shrobe, Waters & Sussman (1979).

The students' databases can be broken down into basic relational patterns. These patterns are reliably associated with particular program structures, allowing us to predict in 80% of the cases precisely what the students' program organization will be. To illustrate this point, consider the following typical (student-generated) database:

```
BURGESS                    PHILBY
 |---ISA--->COMMUNIST       |---ISA--->COMMUNIST
 |---WASAT--->CAMBRIDGE     |---WASAT--->CAMBRIDGE
 |---WORKSFOR--->PHILBY     |---WORKSFOR--->BLUNT

BLUNT
 |---ISA--->COMMUNIST
 |---WASAT--->CAMBRIDGE
 |---WORKSFOR--->THEQUEEN
```

This database exhibits the following patterns:

```
  Transitive-Relation:
        WORKSFOR (BURGESS PHILBY BLUNT THEQUEEN)

  Shared-Successor:
        (ISA COMMUNIST)    (BURGESS PHILBY BLUNT)
        (WASAT CAMBRIDGE)  (BURGESS PHILBY BLUNT)
```

Several items (BURGESS, PHILBY, BLUNT) are present in both the Transitive-Relation and the Shared-Successor lists. One of the program structures typically accompanying such a database structure contains three segments: a CONJUNCTION (COMPLEX-FILTER), SIDE-EFFECT, and GENERATE-NEXT-OBJECT. The Shared-Successors will be used as a FILTER selectively to choose nodes on which a SIDE-EFFECT is perpetrated (e.g. asserting X ISA SPY) and the Transitive-Relation list will be used to GENERATE-NEXT-OBJECT. Since the Shared-Successor list contains two patterns, a COMPLEX-FILTER will almost surely be used: IF x isa communist AND x wasat cambridge THEN ...

But why should a student write a program like this in the first place? Our analysis indicates that students have their own stylised interpretations, or mental models, of the task at hand. For instance, some students think that a program involving two inferences is called for: "If X has done something criminal then he is guilty. And if this is so, then his employer is subject to the same scrutiny, and so on for all employers." Other students feel that only one inference is called for: "Assuming X is guilty, his employer, by association, is guilty also, and so on for all employers."

The observed program structures ought to correspond to students' mental models of the task. Some of these mental models are "appropriate", in that they address the problem as stated, while others introduce certain anomalies which preclude a satisfactory solution. Such "inappropriate" models could actually be artifacts of students "thinking in SOLO" and getting led astray.

In order to test these ideas we have begun studying individual students in depth, collecting videotaped protocols and terminal session transcripts. The first subject began her project session with the clear intention of writing a program involving two inferences. Because of preconceived and inaccurate notions about constraints on the way she was allowed to approach the problem and because of misconceptions arising from her interaction with SOLO, she twice altered her intentions. At the end of the session the

student had a working program for a "one-inference" interpretation of the task described above, a compromise with which she herself was not completely satisfied. All of her programming behavior throughout this session, her various approaches to solving the problem, and the bugs she encountered, fell within the scope of the structures we had identified in the earlier analysis of 160 programs.

Our ability to categorize standard database structures and predict implementation strategies on the basis of those structures means that we can develop tools for assisting students in terms of the way in which they prefer to think about the task at hand. One such tool is described in the next section.

3. Debugging Aids

We have designed and partially implemented a tutorial debugging assistant which attempts to articulate the causes of bugs in terms which are close to the way we believe the students actually think about their own implementations. The bugs dealt with range from domain independent violations of the semantics of SOLO to domain specific errors that can be detected only if knowledge about the task at hand is used.

The debugging assistant uses symbolic evaluation (cf. Ruth, 1976) as a tool for (1) recognizing procedures as parts of a given "library plan", (2) detecting errors of the following types: unreachable program steps, purposeless steps, reference to absent database objects, infinite recursion because of a missing or unsatisfiable termination condition.

In the tutorial situation, a student's goal is to write a program to accomplish some modelling task. The debugging assistant is provided with a prototypical solution in terms of a canonical effect description. The task of the assistant is to recognize a match between the canonical effect description and an effect description derived from the student's own program.

In general these will not match, and the nature of the deviation will enable the assistant to draw the student's attention to shortcomings of his or her program which may be classified in the following way:

- The program will achieve the desired effect only in certain cases. A counter-example outside this range can provoke the student to discover the cause.

- The program would work if missing data or inconsistent entries in the data base were corrected. These corrections can be pointed out directly.

  A particular sub-procedure, if corrected using heuristics about typical errors (e.g. missing indirect link, violation of a program schema), would make the overall program correct. In this case, an appropriate hint can be provided for the student.

  None of the above.

In the last case, the student may initiate a dialogue, requesting help on a particular procedure. During the dialogue the assistant tries to find out the intended effect of that procedure (Eisenstadt & Laubsch, 1980). It does this by categorizing the procedure into one of several programming schemas stored in a language-independent "plan library", using a notation developed by Rich & Shrobe (1978).

Consider the case in which the nearest matching schema is "conjunctive filter and side effect". The assistant examines the deviation between the user's procedure and the stored schema. The following violations of the use of that schema may be recognized: omission of a conjunct, omission of the side effect, wrong (or transposed) arguments in the slots of the schema, or wrong control flow links. The assistant can describe these violations in terms which the student can relate to, since the library plans are themselves based upon schemas known to occur in students' code.

Since the students' procedures depend on their databases, and vice versa, the debugging assistant relies heavily on domain-specific knowledge to deal with alternative ways of formulating a solution to a given problem. Although the students have a great deal of freedom to choose ways of implementing solutions, they typically resort to a few common approaches. The assistant knows about these, and uses these both to make sense of what the students are attempting and to explain why they have gone astray.

4. Conclusions

Our experience with SOLO leads us to believe that a SOLO-like language/environment/curriculum could be of use to a broader group of cognitive science students-- for instance as the basis of a beginners' LISP curriculum oriented entirely towards pattern-matching and assertional data bases. For this to become a reality, it is important to understand precisely what remaining problems students have in this type of environment, and why. Our empirical work is a step in this direction. It has immediate spinoffs in that it provides a foundation for our debugging assistant. The assistant provides students with a tool for attaining their goals, and provides us with a tool for analyzing and describing their behaviour.

REFERENCES

Eisenstadt, M. Artificial intelligence project. Units 3/4 of Cognitive psychology: a third level course. Milton Keynes: Open University Press, 1978.

Eisenstadt, M. & Laubsch, J. Towards an automated debugging assistant for novice programmers. Proceedings of the AISB-80 conference on Artificial Intelligence, Amsterdam, 1980.

Lewis, M. Improving SOLO's user-interface: an empirical study of user behaviour and proposals for cost-effective enhancements to SOLO. Technical report no. 7, Computer Assisted Learning Research Group, The Open University, 1980.

Rich, C. & Shrobe, H. Initial report on a LISP programmer's apprentice. IEEE Transactions on Software Engineering, SE-4:6, 1978.

Ruth, G.R. Intelligent program analysis. Artificial intelligence, 7, 1976.

Shrobe, H., Waters, R., & Sussman, G. A hypothetical monologue illustrating the knowledge underlying program analysis. MIT Artificial Intelligence Laboratory Memo 507, 1979.

Teitelbaum, T., & Reps, T. The Cornell program synthesizer: a syntax-directed programming environment. Technical report 80-421, Department of Computer Science, Cornell University, 1980.

New Tools for Cognitive Science*

Leonard Friedman
Jet Propulsion Laboratory,
California Institute of Technology
Pasadena, CA 91109

Both AI and Cognitive Science must deal with uncertainty much of the time. To cope with this problem, new systems are being developed in AI for representing and propagating subjective belief using semantic nets. In these systems, propagation of uncertainty goes on while logical inferences are drawn. Cognitive scientists may find these methods useful for applications in learning and problem-solving, so this paper will describe the nature of the tools and mention some examples of applications.

There is a long history of attempts by logicians and mathematicians to represent human reasoning more or less realistically. Two basic methods have been employed. One approach is the path of inference, the drawing of conclusions from "givens". The other is the path of likelihood, the estimation of certainty on the basis of experience of some kind. The first theories to be solidly founded have been formal mathematical logic and the theory of probability. Unfortunately, formal logic applies only to a narrow class of situations, and most human reasoning is outside its scope. Similarly, to be applied, probability often demands knowledge not possessed by people. What would be most desirable would be a wedding of inference and likelihood, so that degrees of ignorance could be associated with assertions without requiring unavailable knowledge.

In order to make the contributions of the new methods clearer, we shall describe the nature of the modelling limitations in the older formalisms. Logics such as propositional calculus and first-order predicate calculus demand certainty of belief in the truth or falsity of assertions. In addition, they are monotonic; i.e., they are unable to alter beliefs once established, and also possess no representation of passage of time. One by one these modelling limitatations are being overcome. A variety of non-monotonic logics have been developed which permit the altering of established beliefs. They do this by representing the passage of time in successive "context" layers, each of which is a snapshot of the state of belief in the facts of the universe of discourse. As new evidence is introduced, concomitant shifts of belief are permitted and propagated.

Psychologists long ago proved that ordinary human reasoning is often in disagreement with the dictates of strict probability theory. That theory demands knowledge of probability distributions, gathered in a usually laborious fashion. Situations in which the probability of an event depends conditionally on many other events are computed by using Bayes formula. Bayesian statistics requires a knowledge of a large number of probabilities, not often known to the investigator. On the other hand, humans may reason successfully in situations where they are uncertain and possess no statistical or probabilistic knowledge.

The new methods offer the possibility of modelling some aspects of this type of reasoning. The techniques assume a general knowledge of facts and interrelationships while not requiring detailed statistics. They have been developed by modifying logic on the one hand and introducing parameters that replace probability on the other. We shall not mention the numerous logicians who have contributed to what is called confirmation theory. Zadeh, as early as 1965, began the development of "Fuzzy Logic" (Zadeh '65). Some years later, Shortliffe and Buchanan developed a method of representing degrees of subjective belief or disbelief numerically (Shortliffe and Buchanan '75, Shortliffe '76).

Drawing on the work of the confirmation theorists in logic, Shortliffe and Buchanan found a formulation by which they could fulfill certain criteria established by these workers and at the same time draw "reasonable" inferences with which they associated numerical degrees of certainty. Their work was limited to a narrow class of expressions in the propositional calculus. The implementation was monotonic; i.e., belief in a fact could only grow as new supporting evidence was adduced. Contrary evidence also permitted disbelief to grow when that was appropriate. Evidence for and against a hypothesis was weighed by taking the difference between belief and disbelief. In the MYCIN system, they applied a single mode of inference, confirmation, to medical diagnosis, and called it inexact reasoning. Their significant contribution was to provide a means for making logical inferences based on subjective certainties. Exact formulas were given for the propagation of subjective belief. The formulas depended on an initial assignment of belief-transfer coefficients by a human "expert".

My own work has been concerned with generalizing the methods and refining the formulas to apply to arbitrary expressions of the propositional calculus, employing four rules of inference (ponens, tollens, confirmation, and denial) rather than the single confirmation rule. Also, the implementation is a non-monotonic logic, thus permitting both belief and disbelief to fluctuate according to the evidence. The set of logical rules is called plausible inference and the implementation is named the PI system (Friedman '80a, '81). It is a general logical system, reasoning in either direction, unlike MYCIN which was limited to backward chaining of expressions in a simple form. It also has the ability, on the basis of new evidence, to make its own dynamic assignments of belief-transfer coefficients in certain situations. This ability is essential for learning, as the coefficients are a measure of the relevance of one fact to another. The PI system has been applied to fault diagnosis of a spacecraft (Friedman '80b).

While this line of development was taking place, two mathematicians, Dempster and his pupil Shafer, were independently developing a mathematical theory of evidence (Shafer '76). This theory tackled the problem of representing the degree of ignorance and calculating the likelihood of evidence whether based on objective or subjective considerations. By objective we mean based on formal probability. They showed that measures could be devised in a very general way to

---

* This paper incorporates the results of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract NAS-700, sponsored by the National Aeronautics and Space Administration.

calculate subjective likelihoods. Their formulas had as limiting cases the results of probability theory. Barnett has shown that for certain conditions that apply to our representation, Dempster's general combining formula for the calculation of likelihood or certainty reduces to that employed by Shortliffe and Buchanan, and by myself in plausible inference (Barnett '81). However, Shortliffe and Buchanan's basic formulation was an ad hoc attempt to fit certain logical criteria, so both their rules and those of plausible inference lack a solid mathematical foundation in the computation of belief. Shafer's work shows that the present rules of plausible inference are applicable only for a restricted set of cases, but also provides the information that makes it possible to augment the rules so that it is applicable to most cases of interest, and solidly founded.

For his thesis, a doctoral candidate in AI, John Lowrance, is applying Dempster and Shafer's work to a problem in vision. Recently he and several co-workers have drafted a paper that applies the Dempster and Shafer rules to a different problem (Garvey, Lowrance and Fischler '81). They are estimating the source of a given set of noisy measurements when the origin of those measurements comes from one of a known set of emitters. The measurements are combined to give the degree of support and the uncertainty associated with the evidence for each emitter before that belief is propagated to other assertions via inference.

Such quantities are exactly what is needed in completely automated diagnostic or decision-making inference systems that must deal with uncertainty. The measuring devices would feed degree of belief into assertions linked into a knowledge base, and by plausible inference the knowledge base could draw conclusions about what to do or what went wrong. Garvey, Lowrance and Fischler also point out the possibility of constructing an evidential propositional calculus similar to plausible inference, and suggest coupling the measured estimates based on Dempster's rule with such an inference system.

Summing up, extensions to both formal logic and likelihood theory are converging. The representation and propagation of subjective uncertainty in a knowledge base have been reduced to a set of logical and computational procedures which have propositional calculus as a limiting case. Earlier attempts in cognitive science to model such phenomena include Colby's model of paranoia implemented in PARRY (Colby '73), and Rieger's use of inference to model language understanding (Rieger '76). The new methods have already been applied to a variety of problems in diagnosis, vision analysis, and noisy measurement. Their application to learning as a problem solving activity appears attractive. Further developments in the theory may be possible such as a modification of first order predicate calculus that represents uncertainty.

References

Barnett 81, "Computational Methods for a Mathematical Theory of Evidence", submitted to 7th IJCAI, August 1981.

Colby '73, "Simulations of Belief Systems", Chapter Six, in "Computer Models of Thought and Language" Ed. Schank and Colby, W. H. Freeman and Co., San Francisco, 1973.

Friedman '80a, "Reasoning by Plausible Inference", Proc. 5th Conf. on Automated Deduction, Les Arcs, France, July 1980. Springer-Verlag, Berlin, 1980, pp. 126-142. (No. 87 in the series "Lecture Notes in Computer Science".)

Friedman '80b, "Trouble-Shooting by Plausible Inference", Proc. First Nat. Conf. on AI, Aug. 18-21, 1980, pp 292-294.

Friedman '81, "Extended Plausible Inference", submitted to 7th IJCAI, August 1981.

Garvey, Lowrance, and Fischler '81, "An Inference Technique for Integrating Knowledge from Disparate Sources", submitted to 7th IJCAI, August 1981.

Rieger '76, "An Organization of Knowledge for Problem Solving and Language Comprehension", Artificial Intelligence, Vol. 7, No. 2, Summer 1976.

Shafer '76, "A Mathematical Theory of Evidence", Princeton University Press, Princeton, New Jersey, 1976.

Shortliffe and Buchanan '75, "A Model of Inexact Reasoning in Medicine", Math. Biosci. 23, pp 351-379, 1975.

Shortliffe '76, "Computer-Based Medical Consultations:MYCIN", American Elsevier, New York, 1976, Chapter 4.

Zadeh '65, "Fuzzy Sets", Information and Control 8, pp. 338-353, 1965.

Alf Zimmer

Dept. Psych. - University of Muenster - Fed. Rep. Germany

Cultural Constraints on Cognitive Representation

Starting from Kant's notion of schemata as being rules of synthesis, this paper attempts to develop a unified characterization of innate and culturally mediated schemata. The core concept for this approach is the translation or transformation from one code to another or to some others, which preserve a set of invariants. It is argued that the alleged mechanism of transformation makes possible the development of schemata for complex externally mediated tasks, which are at variance with innate schemata, examples for such tasks are dancing tango or understanding perspective drawings.

Such a notion of schema bears a strong structural similarity to algebraic groups with certain invariant-preserving transformations. This concept can bind together Helmholtz' 'unconscious conclusions' and Hering's explanation of perceptual constancies. It is suggested to apply heuristically the Gestalt principles and laws of perception as invariant-preserving transformations on the content of schemata. By means of an analysis of different tasks the heuristic value of this approach is demonstrated.

Exactly 200 years ago Kant turned around the argument of Berkeley and other Empiricists concerning the truth of abstract ideas, by stating that the a-priory existence of certain rules makes possible the perception of particular instances as such and not vice versa.1) The juxtaposition of a-prioristic and Empiricist views of perception and representation of knowledge was modified further by Helmholtz, who demonstrated that the geometry of the perceptual space is not Euclidean under certain conditions. Therefore the imposed rules are not a-priori but depend on the perceptual task and the cultural and/or phylogenetic development. An even more radical modification is stated in the Sapir-Whorf hypothesis, which pointed out the important intermediate role of language. Despite its apparent relevance for models of human knowledge representation the very role of cultural knowledge in the development and usage of individual knowledge has remained largely undefined.

The comparison of the well-known Necker cube in figure 1, and of a slightly varied form in figure 2 raises immediatly the question, why figure 2 appears to be more 'right'.

Figures 1 & 2

Comparison of two ways to draw a cube

Figure 1

parallel edges of the drawn object are mapped into parallel lines of the same length as in the drawn object

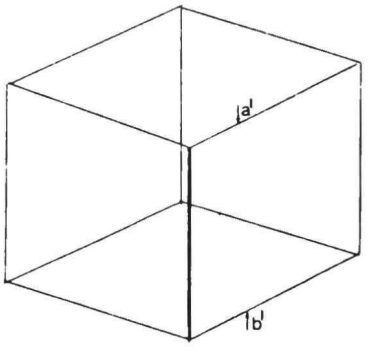
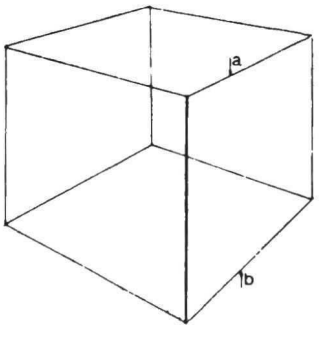


Figure 2

parallel edges are drawn as lines with the same vanishing point and the length of the lines depends on the perspective foreshortening



1) Nietzsche summarized the content of Kant's Critique of Pure Reason in a witticism: "The content is quite simple; Kant showed that the ordinary man is right about perception and that the scientists are wrong. Unfortunately this content is hidden in language, which can only be understood by scientists."

The answer that figure 2 is exactly drawn as we perceive a cube, is not quite convincing; for it provokes the further question, why the art of perspective drawing is a comparatively recent development (about 500 years old), why most people are not able to do it without instruction, and why in cultures not influenced by Western Renaissance art or Chinese art one does not find geometrical perspective. Even if these broader cultural considerations are not taken into account, there remains the puzzling question, why in the context of the cube drawings in figures 1 and 2 the slanted lines a and b in 2 are easily seen as the parallel edges of the right side of the cube, whereas the really parallel lines a' and b' in figure 1 do not elicit this impression as easily.

Various investigations have corroborated the existence of perceptual regularities like the ones of Gestalt psychology; one striking example is the grouping due to the 'same fate' (e. g., parallel curves). If it is assumed that grouping parallel lines is an innate rule for information processing --and it is difficult to see how these lines coulf give rise to the impression of a cube otherwise1) --then the equating of 'straight lines meeting in the same vanishing point' with 'parallel lines on a slanted plane' points out, how cultural constraints can influence the rules, which govern human information processing.

This notion of rules, which allow the perception of spatial forms as such, is due to Kant in his argument against Berkeley's denial of truth in geometry (Critique of Pure Reason). Even if Kant's insisting on these rules as a-priori truths is dismissed nowadays, his development of a schema theory gave rise to conceptualizations in psychology, which opened ways out of the paradoxical situations, into which overconfident reductionism had led psychology. In this context it suffices to reconstruct Kant's idea and some of its applications and further developments in order to point out, what a schema theoretic approach to the introductory problem could look like.

Berkeley had stated that there can be no truth in geometry since theorems of geometry pertain to infinite classes of objects in space (e. g., triangles). According to the empiristic criterion of truth as based on perception (esse est percipi), there cannot exist truth for infinite collections, because that would necessitate the storage of an infinite number of sense impressions (images) for each such class in the human brain. Kant (in 'Critique of Pure Reason') inverts Berkeley's argument by starting from the concept of rules giving rise to an infinite number of images, which can be paired with figures in space: "The schema of the triangle can exist nowhere but in thought and signifies a rule of synthesis of the imagination in respect to its figures in space."1) The connection between the figures in space and their representation in the human brain is made possible through active perception: "... imagination is a necessary ingredient of perception itself." And Kant resolves Berkeley's paradoxical conclusion about the storage of infinitely many images for abstract concepts, by coupling the abstract concept with the active device of the schema, which is able to produce all possible examples for a given concept: "These images can be connected with the concept only by means of the schema to which they belong. In themselves they are never completely at one with the concept."

These ideas have had a strong influence on physiologists in the last century and in the beginning of this century: on Helmholtz and Hering in Germany, on Head in England. The transformations of Kant's schema theory by these scientists in turn influenced Gestalt psychology in continental Europe and Bartlett's conceptualization of psychology in England. Especially the application of this theoretical framework by Head and Bartlett has obscured the distinction between the abstract concept, the active schema, and the ad-hoc produced images. This led to such a vague definition of schema, that this label could be used to denote any steps in the chain of information procession, which seemed to be too complex to be analyzed further. Rumelhart (1980) and others have argued that for the concept of schema to play an important part in cognitive psychology, it is necessary to clarify its definition an d to develop mechanisms to test the feasibility of the concept. But even Rumelhart's notion of 'Schemata as Building Blocks of Cognition' organized in a hierarchical way, seems to have certain weaknesses:

- it necessitates the assumption of primitives in order to avoid infinite regress downwards, while lacking self-evident criteria of demarcation between schemata and primitives
- it does not exclude a possible infinite regress upwards; there always is still another schema necessary to control the top most schema

An alternative strategy to define the concept of schema and at the same time to circumvent some of the mentioned pitfalls, can start by going back to Kant's notion of schema as 'a rule of synthesis', which formed the theoretical framework for Helmholtz', Hering's and Gelb's explanations of constancies in human perception. Helmhotz: "... the law of sufficient reason is really nothing more than the

1) There is an infinite number of two and three dimensional forms which are in agreement with the topologies of figure 1 and 2, which by the way are topologically equivalent.

1) italics by the author

'urge' of our intellect to bring all our perceptions under control." Hering: "... objective knowledge and objective judgment are rendered possible by this constancy," and Gelb: "The idea of invariance, which is an epistemological problem of validity of the foremost importance, has one of its roots, and perhaps the most intuitive one in the psychology of perception."

The search for such invariants and for the related classes of transformations formed the research program of Gestalt psychologists, as can be seen in Köhler's summing up the basic ideas of Ehrenfels: "It is characteristic of phenomenal forms that their specific properties remain unchanged, when the absolute data upon which they rest, undergo certain modification." In Gestalt psychology rules have been developed which pin down invariants and their transformations in perception: the 'laws' of grouping, of saliency (Prägnanz), etc. Unfortunately in the generalizations to more complex forms of human behavior much of the original clarity has been lost. Another limitation of the Gestalt approach is its nativistic orientation, which prevented Gestalt psychologists to realize the interaction between these innate rules and external (e. g., social or cultural) knowledge.

Various analyses (e. g., Uttal, Julesz) have shown how the application of simple filters or auto- and cross-correlation techniques are able to detect regularities (e. g., figures) in apparently random patterns; these results are under certain conditions (types of symmetry, parallel curves, straight lines) the same ones, as one gets with human observers. If one takes the underlying rules, which enable the organism to detect forms in random noise or to discriminate between different forms, as the most primitive building blocks of perception, then it is possible to explain the emergence of schemata. These schemata can be defined as unique combinations of the primitive building blocks under constraints of the organism (e.g., channel capacity, coding capacity) and the environment, upon which these schemata are applied.

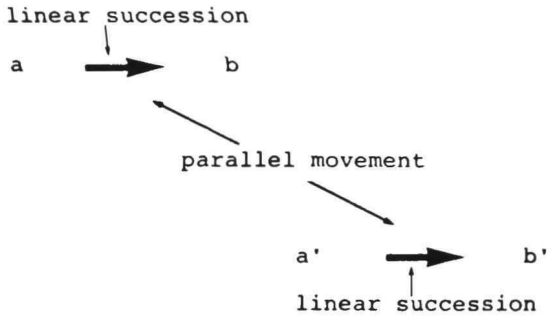
One major objection to this view could be that the invoked rules might be true only for certain modalities; the fact that 'Gestalt-laws' mostly stem from research in visual perception seems to corroborate this critique. What would be the consequence for an assumption of modality specific rules? At first it necessitates further non-analyzable units, which make possible the translation from one sensory code into another, in order to explain highly integrated human activities such as script-writing (integration of kinesthetic, visual and 'symbolic' codes) or dancing, where kinesthetic, vestibular, and acoustic codes influence each other. These activities are not only highly learnable, but they exhibit the mentioned kind of integration from the very beginning (concededly in varying degrees). Furthermore it would be difficult under this assumption to explain synesthetics, the emergence of immediate impressions of similarity between experiences in different modalities (which give rise to self-explaining similes and metaphors) or intermodal rivalry and its resolution. To sum up: The assumption of modality specific rules gives rise to a less parsimonious explanation of inner organismic communication.

This negative argument for modality independent rules can be supported by the demonstration of structurally equivalent rules in different codes: e. g. In grouping perceived objects the rules of linear succession and parallel movements exert strong influence. These very rules are structurally equivalent to the mechanism assumed for the solution of analogical problems by Rumelhart & Abrahamson (1973), as can be seen in figure 3:

If b is linear succession of a, and b' is a linear succession of a', and if the movement from a to b and from a' to b' are parallel, then b' ist the solution of the analogical problem a : b = a' : ?

a, b, a', and b' can stand for verbally stated concepts, for bodily movements, or for acoustical phenomena, and thus can give rise to the resolution of such different tasks as finding the correct verbal expression, as doing the right leg-arm coordination in skiing, or as composing a fugue. It is apparent that in these applications only the set of rules remained invariant, whereas the internal as well as the external constraints changed totally.

Figure 3



perceptual rules underlying analogical problems of type

a : b = a' : ?

On the background of this discussion a schema is a set of primitive rules which are coupled in a unique combination under intra-organismic and extra-organismic constraints for a certain task. In the framework of this task the schema appears as a no-further-analyzable unit and can be considered as a pragmatic primitive. Task hierarchies then imply hierarchies of schema too, thus indicating that the focal point in the perception of the task hierarchy influences, what can be determined as a pragmatic primitive. Form this point of view the function of the schema can be assumed as twofold:

#1 the generation of possible constellations (e. g., strings of symbols, sequences of movements, figures) and their comparison with perceived constellations

#2 the extrapolation of future states of the organism and their comparison with perceived states.

Whereas the former allows identification and classification of information and events, the latter makes self-regulated goal-directed behavior possible. A combination of both functions underlies behavior, which is regulated by external feedback.

The influence of cultural constraints on the functioning of schemata can be observed best in instances, where the rules of the cultural constraints interfere with the rules of an either inborn schema (e. g., upright walking of humans with a characteristic arm-leg coordination) or an individually developed schema (e. g., certain motor patterns which make up the invidual characteristics of script, drawing, or gesticulation). Such an interference is mostly possible, if the existing schema can produce particulars (in Cassirer's sense) in agreement with the cultural constraints in the limiting case but not in the general case. This situation can be clarified by two examples from different domains.

The cultural constraints on a mode of argumentation typical for the Western culture can be found in the rules of syllogistic reasoning. Psychological analyses of this task since Wertheimer have demonstrated, that formally equivalent problems lead to totally different patterns of successes and errors, and that framing, which does not alter the logical structure, influences the probability of success in a highly predictable way. If the rules of perception are taken as primitives for a schema of argumentation, then these rules, especially 'neighborhood', 'succession', 'same fate', and 'parallel curves', can be taken as the underlying cause for Johnson-Laird's figurative effect as well as for the superiority of concrete problems over abstract problems of the same logical structure.

The learning of movements in dancing is another example for the interference of cultural constraints with inborn schemata. Physiological as well as psychological analyses of human motor behavior have demonstrated that its apparent smoothness is due to CNS-controlled 'pre-enervations', which allow a very fast sequence of the same or different movements, e. g. in walking (same), or jumping (different). A necessary precondition of these 'pre-enervations' seems to be, that one movement follows the ather in 'a natural way'. If the perceptual rules are assumed to characterize this 'natural way', then it becomes plausible, why tango seems to be one of the most difficult dances to learn. The initial sequence of steps makes use of the pattern of normal walking. Part of the walking schema is stopping, but in tango instead of a simple stop with subsequent change of direction a kind of oscillation of the body is preseribed, which is only found in this type of dancing. Thus the invariants of tango are incompatible with the schemata of walking as well as of other kinds of dancing. The obvious fact, that this movement can be learned nevertheless, may be made possible by the translation of the rhythmic invariants contained in the melody, into rhythmic patterns of bodily movements. The fact, that dancing tango without heard or imagined music is nearly impossible, makes this conclusion plausible.

In order to resolve all the questions and vicious circles in relation to schema theory and its interdependence with cultural constraints a more detailed and experimentally corroborated theory is necessary than this admittedly uneven and loopholed sketch of an alternative starting point can be. Nevertheless it is hoped that at least it stimulates clarifying discussions about this central concept of cognitive psychology.

These ideas have gained much from David Rumelhart's lectures and discussions with Jennifer Freyd and Regine May. The remaining lack of clarity must totally be contributed to the author

Some thoughts on logic, language, mind and reality

Rachel Joffe Falmagne
Clark University

This discussion will be predicated upon the notion that adult human cognition includes an internalized system of logical relationships and principles, in other words, that logic has some psychological reality in that at least a fragment of it is included in human competence.

To avoid some likely misconstruals, it seems a good idea to state here what this discussion is not. It does not examine whether standard logic, as a complete system, adequately represents human reasoning - a question that has preoccupied psychologists and some philosophers for some time (e.g. Cohen, in press) - nor does it attempt to determine whether an alternative logical system would be more suitable for this purpose as some psychologists have proposed. Nor does it take a stance in the debate regarding the adequate logical system in terms of logicians' and philosophers' criteria (ostensibly distinct from those of the psychologist). Rather it is assumed that some logic, possibly incomplete, is part of human knowledge; which form it has does not matter for the sake of the present discussion.

Given this premise, the aims of this paper are (i) to provide a first step towards integration of some issues in philosophy of logic with analogous issues concerning human logical cognition, in particular to examine the relationships between logic and language and logic and reality from those two vantage points; (ii) to address some general questions concerning the development of logical cognition within that perspective. In relation to this second aim, I will argue, more generally, for the necessity of articulating the epistemological and developmental foundations of the cognitive functions or knowledge structures that we are investigating in cognitive science, if our account of those functions and structures is to have any genuine substance.

The relationships between logic, language, and reality are notoriously controversial among philosophers and linguists. This paper cannot begin to give an adequate account of the debate and of its philosophical and conceptual ramifications, but as an oversimplified approximation, the issues that will concern us here, have to do with the nature of propositions as, alternatively, empirical linguistic, or subjective entities, and, correspondingly, whether logic is more properly ssen as formalizing necessary relations in the empirical world, or analytic relations within natural language. What follows is a schematic survey of a few alternative positions, limited to highlighting the relevant contrasts. Thus, for example, Carnap (1951) and Katz (1972), on different grounds, put forth a view of logical relations as linguistic relations, and characterize logical truths in terms of linguistic structures, taking sentences rather than abstract propositions as primitives. Putnam (1971), on the other hand "finds something ridiculous in the theory that logic is about sentences" (ibid, p.6) and, in this article and others, defends a realist position with respect to the entities subsumed within logic (classes, properties, etc...), and, presumably, with respect to logical truths. Frege (1918/1956) similarly "assigns to logic the task of discovering the laws of truth, not of assertion or thought" (ibid p.289), thereby excluding both language and mind from the foundations of logic. Quine, on the other hand, argues against the distinction between analytic and synthetic truths (Quine, 1953) and, relatedly, sees logic as one part of a "whole interlocked scientific system", an integral part of any scientific prediction, and therefore ultimately revisable on the basis of empirical evidence (Quine, 1970); the close relations between logic and language are acknowledged but the foundations of logic are nonlinguistic (and, additionally, empirical). Interestingly, a somewhat analogous relationship between logic and other forms of knowledge, is present in Wittgenstein (1922/1961), despite the radical divergence of those two philosophers on most other matters: starting with the premise that the world is the totality of facts, the early Wittgenstein's thesis is that we picture facts to ourselves, that a picture is a model of reality, and has in common with it its pictorial form. A logical picture of reality is one which only has logical form in common with the reality it depicts; therefore, every picture is at the same time a logical picture. Thus, on this account logical relations do not have a status distinct from that of other modes of representing reality: they are grounded directly in the structure of events, as other modes of representation are, although they are more basic and of wider applicability.

Analogous questions arise with respect to logical cognition and will be discussed here. They articulate with the general context of the philosophical issues just mentioned in two ways. First, examining human logical cognition entails assumptions about the nature of logical knowledge. Second, examining logical development entails assumptions about the sources from which logical knowledge is derived.

Regarding the former question, the specific issue is whether logical knowledge primarily consists of knowledge about the structure of language or of knowledge about the structure of events in the empirical world. The psychological angle on this issue highlights considerations of a somewhat different kind than those of the philosopher. The concern is not so much about the foundations of logic but rather about the way in which logical knowledge articulates with the other components of an individual's knowledge structure. In an important sense, logical knowledge is knowledge about both language and the empirical world to the extent that language itself is semantically grounded in the empirical world to which parts of it refer. However, cognitively in terms of the organization of the mind - logical knowledge may or may not hinge on linguistic knowledge. Thus, for example, regardless of the fate of the philosophical debate about whether the notion of analytic truth is well-founded, (e.g. Quine, 1953; Putnam, 1962), one may ask whether, cognitively, people can discover new logical truths via linguistic structures exclusively - a cognitive analogue of the "analytic" situation. This question is seen most clearly when put in a developmental light, as will be done presently.

The second question raised previously concerned the epistemology and development of logical cognition. It will be assumed here that logical knowledge is derived constructively both from linguistic and nonlinguistic sources.

Regarding the linguistic foundation first, the assumption is that logical knowledge is derived both from awareness of the structure of language itself and from the correspondences between linguistically expressed propositions and empirical states of affairs, with the latter source of knowledge ensuring that the resulting logical system remains semantically sound and internally consistent (though not necessarily complete).

More specifically, the initial comprehension of "logical" relations is certainly semantically based and contextually restricted, as has been shown in the language development literature for other kinds of relations. However, it seems compelling to assume that further elaboration of these logical relations involves a process of abstraction from their initial content-bound meaning and an elaboration of their linguistic properties. The development of negation may be a paradigm example of this microdevelopmental process. Negation in the early stages of language development appears to signify disappearance or nonexistence, and it only subsequently emerges as a propositional operator in children of 2-3 years of age (Pea, 1980). However, by age 5 or 6, it seems hardly questionable that the logical properties of negation are mastered by the child at a fairly abstract level, in the sense that the child knows that if a statement p is true, its negation is false across a wide range of contents and presumably on inferential grounds. Thus, negation initially appears to refer directly to the events or objects themselves, and its meaning is grounded in direct verification of the presence or absence of these objects. Further developments, however, are of a more "syntactic" kind, though presumably retaining the initial meaning as their semantic foundation. With regard to this later development, it is useful to remember that logic and syntax as formal systems, have a highly similar status with respect to natural language. Although the details of the parallel, its limitations, and the resulting issues are beyond the scope of this paper, it is enough to note that the two systems are alternative formalizations of natural language and that both logic and syntax interface with semantics, in ways that are partly similar. If one recognizes this parallel, it is then natural to look at both syntactic and logical development as a process of gradual structuring of the linguistic environment. Furthermore, it is natural to speculate that, in a way somewhat similar to the way in which the child learns to structure her/his linguistic environment syntactically (presumably by exploiting the interconnections between the syntactic, semantic and contextual aspects of language) s(he) may also be assumed to structure the linguistic environment in terms of what statements can be legitimately derived from what other statements, and under what conditions. Thus, what is suggested here is a process of abstraction of logical forms from content specific instances embodying this form. If, in addition, as some have proposed, logical and grammatical structures are in close correspondence, the acquisition of grammar and of logical forms would proceed concurrently in part.

Turning now to the nonlinguistic sources of logical development, two positions are possible. Piaget takes the most radical option in locating the foundations of logical structures in the systems of actions of the child upon the world and in positing that logical relations are constructed by reflective abstraction on the properties of the world as apprehended by the child's actions (e.g. Piaget and Inhelder, 1969). This is a radically nonlinguistic account. A different account can be given, based on the notion that, except for the very initial period of cognitive development, the events in the world and the relationships between these events can be represented mentally in terms of propositions, so that e.g., valid patterns of inference can be abstracted from the structure of events in the world via propositional encoding. An example will help concretize this notion: in observing the functioning of an electrical circuit possessing implicative properties (turning on either one of two switches, S_1 or S_2, causes the same light to go on), a person may observe that, when the light is on, one does not know which switch has been used - a typical indeterminate conditional inference. Coming back to the earlier part of this discussion, if logic is seen as formalizing necessary relations in the empirical world, observations of the kind just sketched may provide a direct, (ostensibly) nonlinguistic foundation for such logical knowledge.

So far, abstraction of logical knowledge from linguistic and nonlinguistic sources, has been discussed in general terms. A mechanism implementing this conception will be outlined, extending notions discussed in Falmagne (1980), in which various modes of representation of linguistic and nonlinguistic information are assumed to be possible (the formal mode being one of them, and mental models as proposed by Johnson-Laird (1980) being another) and in which functional and developmental relationships between those modes of representation are described. This conception is somewhat parallel, in a mentalistic way, to the early Wittgensteinian notions discussed earlier, and the way in which perhaps logical forms are abstracted from the structure of mental models together with the operations carried out on them, will be discussed.

The preceding discussion should not be mistaken as reflecting an empiricist epistemology. It seems clear to me that a strictly empiricist account of logical development and a strictly rationalist account are equally untenable and, furthermore, intellectually unappealing. An assumption that seems more apt, both on empirical and philosophical grounds, is that natural logic is both constrained and made possible by fundamental properties of the mind - minimally by fundamental cognitive ways of processing experience. What should be invoked on the "constraint" side is far from being clear at this point. Some proposals whose relevance to the present issue needs to be examined or developed are notions of natural connectives, (Osherson, 1977) notions of conceptual naturalness, and, with some qualifications, some linguists' quest for linguistic universals.

On the "positive" side, regarding those properties of the mind that make natural logic possible, one of these basic cognitive functions is the human capacity for abstraction, which provides the mechanism for emergent discontinuities in modes of thinking and of processing language in the course of development or of learning (those discontinuities which radical empiricism is poorly equipped to account for). In the same way as this capacity permits the child to acquire a linguistic medium which stands in a symbolic relation to the referent world, and, later on, to undergo the formal structuring underlying advanced syntactic development, perhaps it permits for logical forms and logical truths themselves, to be abstracted from language (and nonlinguistic experience), as

has been proposed here. Thus, the program as I see it is to understand the interplay between learning mechanisms, and the cognitive constraints and possible a priori dispositions within which learning operates.

References

Carnap, R. The logical syntax of language. Paterson, N.J.: Littlefield, 1959 (first published in 1937).

Cohen, L.J. Can human irrationality be experimentally demonstrated? Behavioral and Brain Sciences, in press.

Falmagne, R.J. The development of logical competence: a psycholinguistic perspective. In Kluwe, R. & Spada, H. (Eds.), Developmental models of thinking. Academic Press, 1980, pp.171-197.

Frege, G. The thought: a logical inquiry. Mind, 1956, 65, 289-311. (First written in 1918)

Johnson-Laird, P.N. Mental models in cognitive science. Cognitive Science, 1980, 4, 71-115.

Katz, J.J. Semantic theory. Harper and Row, 1972.

Osherson, D. Natural connectives: a Chomskyan approach. Journal of Mathematical Psychology, 1977, 16, 1-29.

Pea, R. Development of negation in early child language. In D. Olson (Ed.), The social foundations of language and thought. Norton, 1980.

Piaget, J. & Inhelder, B. The early growth of logic in the child. Routledge and Kegan, 1964.

Putnam, H. Philosophy of logic. Harper and Row, 1971.

Putnam, H. The analytic and the synthetic. In Feigl, H. & Maxwell, G. (Eds.), Minnesota Studies in the Philosophy of Science, III. University of Minnesota Press, 1962.

Quine, W.V. Philosophy of logic. Prentice-Hall, 1970.

Quine, W.V. Two dogmas of empiricism. In W.V.O. Quine, From a logical point of view. Harvard University Press, 1953, pp.20-46.

Wittgenstein, L. Tractatus logico-philosophicus. Routledge and Kegan, 1961. (First English edition in 1922)

Time and Cognition: The Domestication of the Maya Mind

Hugh Gladwin
University of California, Irvine

In his provocative book The Domestication of the Savage Mind Jack Goody argues that the advent of writing has profoundly altered the growth and structure of human knowledge, and beyond that, human cognition itself:

> Writing is critical not simply because it preserves speech over time and space, but because it transforms speech, by abstracting its components, by assisting backward scanning, so that communication by eye creates a different cognitive potentiality for human beings than communication by word of mouth. (Goody 1977:128)

Goody argues that the making of written tables and lists has expanded the potential of cognitive activities. The list, he says,

> is connected in a direct and intimate way with cognitive processes. For the making of lists, actual or figurative, that I have called shopping lists is part of the more general process of planning human action. . . . It is not so much the making of plans, the use of symbolic thought, as the externalizing and communication of those plans, transactions in symbolism, that are the marks of man. . . . [list making] represents one aspect of the process of decontextualization (or better 'recontextualization') that is intrinsic to writing, not merely as an external activity but as an internal one as well. To put the matter in another way, writing enables you to talk freely about your thoughts. (p. 159)

The book concludes that writing

> encourages special forms of linguistic activity associated with developments in particular kinds of problem raising and problem solving, in which the list, the formula, and the table played a seminal part. If we wish to speak of the 'savage mind', these were some of the instruments of its domestication. (p. 162).

In this paper I suggest that Goody's argument fits well with a view of cognition which was dominant in the 1960's. But the last decade's developments in cognitive theory have made Goody's conclusions about the effect of writing on human cognition less tenable (of necessity my review here is cryptic; see Colby, Fernandez and Kronenfeld's [1981] coverage of the same ground from a slightly different viewpoint). I will illustrate these points with a brief example from well known facts about Maya arithmetic and chronological reckoning.

First, I should note that there are many of Goody's arguments that I find very convincing. Surely writing has facilitated an explosive growth in human knowledge stored in libraries, data banks, and other places. This growth in stored knowledge, along with the growth in size and role differentiation of modern society, makes learning even the small fraction of knowledge one must have to be a competent member of society a staggering task (D'Andrade 1980:2-3). And writing has changed many activities, such as most of the arts: the act of composing music, and of writing poetry and plays is now usually separate from the performance. This is so much so that Aaron Copeland noted recently that, as a composer, conducting for him was a great joy -- "... it might not be as good as somebody else's interpretation, but nevertheless it's the way you thought of it". And the performance of performances, as with the performances of many tasks, has changed radically with the advent of writing. It is hard to take exception to Goody's arguments on these topics.

It is Goody's extension of the argument to cognition itself that I want to question. His argument is very subtle: writing reinforces the cognitive facility to represent concepts by symbols (words) which can be moved around into lists, tables, and formulae. In purely oral speech one is always in the "flow of speech", and rarely conscious of individual words (Goody 1977:115). Writing, on the other hand, allows one to reflect on concepts, operate logically on them, and categorize them.

This view of Goody's is congenial with the view of language and cognition which was dominant in the 1960's. In anthropology frames which could be elicited in speech were used to generate lists, tables, and formulae which were thought to accurately reflect cognition. If this activity is a fundamental part of cognition, then surely Goody is right in thinking that writing reinforces an important type of cognitive activity. In anthropology and in transformational linguistics meaning was seen as being composed of the semantics of individual words (often representable in feature notation) plus the logical functions operating on the different parts of the sentence generated by the syntax. It was a view of language and cognition which was very precise and operational. Even anthropologists occasionally used the expression "machine translation" to express the ultimate goals of cognitive science. Again, if this is cognition, writing should train and expand it.

In the 1970's another view of the relationship between language and cognition asserted itself. This position was not new; F. C. Bartlett stated it in 1932, and it has always had advocates in some areas of philosophy. In essence this view argues that the main "work" in communication involves memory structures (schemas) often loosely connected with language. Semantics is thus demoted to the accessing of memory schemas. As I once put it,

> In schema theory the job of semantics is not to completely define words, but rather to show how words are related to memory schemas in the interpretation of sentences. When two people are talking they use arbitrary sound signals, words in sentences, and there must be some basis for agreement between them on what knowledge the words and sentences refer to. Semantics is the basis for this agreement. (Gladwin 1972, 3-5)

Roger Schank's first presentation of the conceptual dependency model of processes underlying the production and interpretation of conversation, as well as the later work by him and his colleagues, (Schank 1972, Schank and Abelson 1977) also posited a much looser view of the relationship between phonology/syntax on one hand and conceptual structures (memory) on the other. A branch of transformational linguistics, generative semantics, grew ever more elaborate semantic trees, which began to look more like models of cognitive process than semantic operations closely tied to syntactic and phonological rules. Few of these cognitive researchers deny the brilliant achievements of Chomsky and his followers in syntax and phonology, but they argue that the syntactic and phonological structures of speech are generated and interpreted almost automatically, outside the conscious attention of speaker or hearer. The thread of the conversation is carried by conceptual structures which seem unlike either syntactic trees or componential paradigms.

Recent work further emphasizes the focus on conceptual structures. Lakoff and Johnson (1980), argue that metaphor, a cognitive analogy, is fundamentally important in the production and interpretation of speech and text; Quinn (1980) has employed the analysis of metaphor in understanding the conversations of people about their marriages. Most current work on folk tales emphasizes conceptual rather than syntactic structures (Colby 1981). Brown and Suchman (1981) have argued that the powerful concepts underlying skilled technical behavior are based on qualitative comparison with devices which are known to work in ways similar to the device involved in the task to be performed. Precise quantitative calculation and logical deduction seem relatively less important than had been thought.

The Brown and Suchman argument has much in common with other work on cognitive structures underlying highly skilled behavior, work which departs markedly from the logic and language-based model of skilled, intelligent thought. It was once widely assumed that chess masters, for example, could operate logically and deductively on the symbols and tokens of chess to evaluate many moves ahead to the consequences of current possible moves. But Chase and Simon (1973) found that the first moves master players attend to are usually the best moves: "Masters invariably explore strong moves, whereas weak players spend considerable time analyzing the consequences of bad moves. The best move, or at least a very good one, just seems to come to the top of a master's list of plausible moves for analysis" (1973:216). They also found that the memory context in which best moves are "recognized" is associated with known board configurations; when presented with randomly generated board configurations masters did much worse than when dealing with configurations likely to occur in actual play.

The Chase and Simon study illustrates two *aspects of cognitive processes underlying highly skilled behavior which have recently been noted. First, the memories recalled are highly dependent on task environments (e.g. board configurations).* Work by members of the Adult Math Skills Project at U. C. Irvine (1979, 1980, 1981, Lave 1981) has emphasized the interaction of task environment and skill knowledge, interaction resulting in a notion of "situational memory" closely analogous to the memory of board configurations in the Chase and Simon study. The Brown and Suchman study emphasizes analogies between devices engineers (and copier operators) know and ones they are trying to figure out. Devices become well known in a given task environment. Second, the process of figuring out what to do in skilled behavior seems to work much more like a recognition task than a deductive task. One "sees" the situation of the task, and "recognizes" what to do, like the chess masters see the board and recognize what to do. Much of the actual problem-solving is accomplished preattentively and before the setting up of what Newell and Simon (1972) would call the "problem space": the problem that people attentively consider and can verbally report on.

At this point in my paper some readers are bound to object that cognitive processes underlying skilled problem solving may seem like recognition, and may seem to be environmentally situated, but a model of the competence required to perform the task need not be concerned with where a task is situated, or whether its performance is consciously attended to or not. This objection may or may not be valid (it's not the purpose of this paper to debate it). My argument is rather that attention and performance greatly affect learning, and in the long run would affect the cumulative expansion of cognitive ability that Goody argues writing brings about. In other words, it's hard to learn something you don't understand or are not aware of. The research reported on here argues that understanding is based largely on conceptual operations, like metaphor and qualitative comparison. And awareness of a task is usually of a task in its environment. It thus seems to me unlikely the "recontextualization" of task instructions, for example, to printed formulae or tables, will in and of itself improve learning or cognitive skills in general. It is less likely to facilitate learning if the task context is completely removed from the written dscription of the task itself. Furthermore, the task is almost certainly unlearnable if the conceptual operations fundamental to its performance are not presented.

The argument can be restated in terms of a distinction made by Roy D'Andrade at the Cognitive Science Conference last year in New Haven. He makes a distinction between what he calls "content based" abstraction and "formal language" abstraction. Content based abstraction is abstraction situated in one context. He illustrates content based abstraction with a chess example similar to the one from Chase and Simon cited above. Formal language abstraction "involves recoding the problem into a different symbol system" (1980:13). On the face of it, formal language abstraction appears like the 60's notion of semantics and cognition discussed above. It is certainly close to Goody's notion of what would be encouraged by the development of writing. But the "70's" position would take exception to D'Andrade's implicit assertion that *only formal language abstraction can be recoded to a different semantic domain. If we take the view that semantics includes only pointers to and from* memory structures, the recoding will take place in memory structures, not in semantics or formal language. Metaphors and Brown and Suchman's

qualitative reasoning are much more likely candidates for memory devices which permit "abstraction" from one domain to another.

I am thus arguing that we are most of the time more like Goody's savages than domesticated people (excepting, of course, logicians and poets). Can the written word help a savage? It can, but only if it allows easy translation in terms of the powerful memory devices needed to perform a task. This is why, as D'Andrade notes, humans often have difficulty learning procedures which require formal language abstraction. In the Adult Math Skills Project we have found that important among of the powerful memory devices for performing measurement calculation and estimation are highly overlearned structures and operations in which the perception of a measure is associated with a quantity. We have called these "canonical units". An example would be knowing that a football field is as big as an acre. Unfortunately British/U.S. measure, while it does access canonical units, does not usually translate well. Most Americans do not know how many feet are on the side of an acre, or how many acres are in a square mile. The corresponding facts are probably known by a much larger proportion of the people who use hectares instead of acres. My argument, then, is that the use of the metric system might very well improve its user's cognitive procedures for spatial calculation and estimation. But I doubt if the ability to write down numbers in and of itself improves the ability to calculate.

There is one area (besides logic and poetry) where formal language abstraction and writing per se is important. That is the environment where formal abstraction and deduction is required whether or not it facilitates learning: school. But that is a subject for another paper.

I will conclude with a comparison of the Western and the Maya systems for calculating dates and elapsed time between dates. Both systems permit the generation of lists, tables, and formulae linking dates and events, operating calendars, etc. But the Maya system appears much more likely to have facilitated the "domestication" of the cognition of its users than the Western.

It should be first noted that the Maya were not so concerned as we are to be able to calculate to a given point in the solar year. In our system most everyone knows that January 1st falls at the same time in the solar year (i.e. in the same part of winter in the U.S.). They were more concerned with "translatability" in the sense that I have used it here. That is, they wanted to have the units of the calendar correspond to canonical units of time. They also wanted the units of time to correspond to basic arithmetic operations. An example of correspondence with arithmetic operations is the metric system of measure, in which most measures are multiples of ten, corresponding to a base ten arithmetic. For the Maya the 360 day year was sufficiently close to the solar year to serve as a canonical unit, and 360 translated both into the Maya arithmetic system and the calculations astronomers wished to perform.

Maya arithmetic is most commonly written in a bar and dot notation. Dots are units and bars are marks for tallying at five. Tallying at five is important for a commercially useful arithmetic since five is within the subitizing range (Klahr 1973, Adult Math Skills Project 1979). Given the growing appreciation of the importance of trade in Maya history (Rathje 1971), we can understand why a tally at five system was very useful. The abacus is another commercially used example of a system which tallies at five. An "integer" in Maya arithmetic is composed of a combination of bars and dots up to 19. This is followed by a shell-like figure for zero. The system is base 20, and the "digits" ("vigits"?) are usually written vertically.

Only a slight modification is then needed to bring the system into correspondence with the canonical units of time. The third "digit" is base 18 rather than base twenty, giving a unit of 360 days (the tun). The calendar thus has the following units:

baktun
400 "years"

katun
20 "years"

tun
one "year", 18 "months", 360 "days"

kin
one "month", 20 "days"

uinal
one "day"

What the metric system does for spatial and weight measure, the Mayan system does for time. Given two dates in the Western calender, on the other hand, it is a tedious task to figure how many days elapsed between two dates. Most people have to resort to counting. But in the Maya system calculation of the interval between two dates is done almost as easily as a user of the metric system can find the difference between 238 cm. and 5.126 m. More information on Maya arithmetic and calendrics can be found in Thompson 1960, Marcus 1976, Aveni 1976, and in the delightful but, alas, out of print book by George I. Sanchez (1961).

I can thus conclude that both Western and Mayan calendars heavily employ writing, list making, and tabulation. But I would argue that the Mayan system is a powerful amplifier of chronological cognition, while the Western calendar is not. Goody is right in seeing writing as "domesticating" the savage mind. But he is wrong in thinking the effect is global; it only works when the writing system translates powerful memory processes well.

REFERENCES CITED

Adult Math Skills Project
1979 Hugh Gladwin. Memory Models from Psychology, Anthropology and AI: Are Any Useful in Understanding Adult Mathematical Problem Solving. Paper presented at the Riverside Conference on Cognition.

1980 Michael Murtaugh, Katherine Faust, and Olivia de la Rocha. Everyday Problem-Solving Events. Paper presented at the American Anthropological Association Meetings, Washington, D.C.

1981 Olivia de la Rocha, Michael Murtaugh and Jean Lave. A Conceptual Framework for Locating Cognitive Process in Daily Life. Presentation at Society for Research in Child Development Workshop, Laguna Beach, California.

Aveni, Anthony F.
1978 Old World and New World Naked-Eye Astronomy. Technology Review 81:60-72.

Bartlett, Frederic C.
1932 Remembering: A Study in Experimental and Social Psychology. Cambridge University Press.

Brown, John Seely and Lucy Suchman
1981 Presentation at Workshop on the Social Context of the Development of Everyday Cognitive Skills, Society for Research in Child Development, Laguna Beach, California.

Chase, William and Herbert Simon
1973 The Mind's Eye in Chess. In Visual Information Processing. W. G. Chase, Editor. pp. 215-281. New York: Academic Press.

Colby, Benjamin N. and Lore M. Colby
1981 The Day-Keeper: the Life and Discourse of an Ixil Diviner. Cambridge: Harvard University Press.

Colby, Benjamin N., James W. Fernandez, and David Kronenfeld
1981 Toward a Convergence of Cognitive and Symbolic Anthropology. American Ethnologist 8.3.

D'Andrade, Roy G.
1980 The Cultural Part of Cognition. Address given to the 2nd Annual Cognitive Science Conference, New Haven.

Gladwin, Hugh
1972 Semantics, Schemata and Kinship. Paper presented at Mathematics in the Social Sciences Board Conference on Semantic Models in Kinship. Riverside, California.

Gladwin, Hugh and Michael Murtaugh
1980 The Attentive-Preattentive Distinction in Agricultural Decision Making. In Agricultural Decision Making: Anthropological Contributions to Rural Development. Peggy Barlett, Editor. pp. 115-135. New York: Academic Press.

Goody, Jack
1977 The Domestication of the Savage Mind. Cambridge University Press.

Klahr, David
1973 Quantification Processes (pp. 3-34) and A Production System for Counting, Subitizing and Adding (pp. 527-546). In Visual Information Processing. W. G. Chase, editor. New York: Academic Press.

Lakoff, George and Mark N. Johnson
1980 Metaphors We Live By. Chicago: University of Chicago Press.

Lave, Jean
1981 Tailored Learning: Education and Cognitive Skills Among Tribal Craftsmen in West Africa. ms., U.C. Irvine.

Marcus, Joyce
1976 The Origins of Mesoamerican Writing. Annual Reviews of Anthropology 5:35-67.

Newell, Alan and Herbert A. Simon
1972 Human Problem Solving. Englewood Cliffs, New Jersey: Prentice-Hall.

Quinn, Naomi
1980 Committment in American Marriage: the Conceptual Structure of a Key Word. Paper presented at the American Anthropological Association Meetings, Washington, D.C.

Rathje, William L.
1971 The Origin and Development of the Classic Lowland Maya. American Antiquity 36: 275-285.

Sanchez, George I.
1961 Arithmetic In Maya. Austin, Texas, privately printed.

Schank, Roger C.
1972 Conceptual Dependency: A Theory of Natural Language Understanding. Cognitive Psychology 3:552-631.

Schank, Roger C., and Robert Abelson
1977 Scripts, Plans, Goals, and Understanding. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Thompson, J. Eric S.
1960 Maya Hieroglyphic Writing: an Introduction. Norman: University of Oklahoma Press.

Why some modified class inclusion tasks are easy for young children: A process model for finding referents of labels in arrays

Adele A. Abrahamsen
New School for Social Research

Inhelder & Piaget's standard class inclusion problem involves the presentation of an array, e.g., five toy horses and three toy dogs, and the posing of a question concerning the larger subordinate class (A_1) and the superordinate class (A): "Are there more horses or more animals?" Children younger than about eight years usually say "More horses."

Where do they go wrong? Wilkinson (1976) has suggested the use of a strategy that forbids double-counting. He outlines a SCAN procedure that young children apply only once (avoiding double-counting), and older children apply twice.

Despite the intuitive appeal of this model, there is evidence that children have difficulties correctly assigning referents in certain related tasks using nonincluded sets, for which double-counting is not at issue (McGarrigle, Grieve & Hughes, 1978; Isen, Riley, Tucker & Trabasso, 1975). Conversely, Markman (1973) showed that children perform well on included sets if the more inclusive set is a collection, e.g., "family," rather than a class. Finally, Trabasso, Isen, Dolecki, McLanahan, Riley & Tucker (1978) reviewed a number of modified class inclusion tasks for which performance is modestly or dramatically better than for the standard task.

Though Trabasso et al. interpreted the findings in terms of eight component processes in class inclusion, they emphasized that the standard task encourages children to identify the smaller subordinate class as the referent of the superordinate term, whereas many of the modified tasks make the superordinate reference unambiguous. In this paper I expand on that insight by outlining two simple process models for finding the referents of labels in arrays, and showing that these models can account for younger and older children's performance on standard tasks and on those modified tasks which elicit dramatic improvements.

Process A (for younger children)

I propose that young children proceed as follows for class inclusion and related problems:

1. They exhaustively partition the array of objects into mutually exclusive sets. I assume that they tend to maximize similarity within sets and minimize similarity between sets, hence favoring small sets, but I do not model this procedure.
2. They seek referents for labels by conducting a self-terminating search of the sets.
3. They permit the referent sets of two labels to intersect (i.e., double-counting), but they cannot be identical; a search that would otherwise be terminated will continue to avoid this.

Figures 1 and 2 show modified flow charts of the two most important procedures of Process A. These are incomplete, but are detailed enough to address the gross empirical data, and are similar in format to Wilkinson's (1976) flow charts to facilitate comparison.

It is assumed that the child has constructed two lists: a list S* of the sets in the array and a list L* of verbal labels. Any relevant noun phrase in the class inclusion question or in the immediate verbal context is included in L*. The procedure LINK acts on these lists, linking labels to sets, and keeping track of the linkages by constructing a third list of labels and their referents: (L,REF)* LINK processes every label in turn, getting its definition and calling the procedure SETSEARCH to find a matching set description. Once a set is found to have a matching description (by the procedure MATCH, not detailed here), the search terminates unless the result is a reference already "claimed" by another label, as determined by the procedure COMPARE. Whenever a set is "claimed" by a label, it is marked as used (#) and moved to the end of the list S*. After all labels have been processed, LINK uses LASTSEARCH to try once again to make a linkage for any remaining unclaimed sets. This is not detailed here, but for each such set a self-terminating search of the labels is performed.

Process B (for older children)

Process B conducts an exhaustive search of the sets for each label, and permits two labels to have identical referent sets. Intermediate processes incorporating only one of these two changes are of course possible. The changes primarily affect SETSEARCH (see Figure 3), but the exhaustive search feature makes the check for leftover sets in LINK unnecessary. Process B obtains the correct answer to standard class inclusion problems, unless the labels are defined too narrowly to match atypical sets.

Evidence

Figures 3 and 4 trace the highlights of Processes A and B, respectively, as applied to several versions of class inclusion tasks. The standard task is followed by tasks in which both subordinate labels are made salient (Ahr & Youniss, 1970; Winer, 1974) and in which several subordinate classes appear in the array (McLanahan, 1976); these versions are easy for young children. Last, three conditions involving atypical classes are shown (Carson & Abrahamson, 1976); these are difficult even for some older children.

With minor modification, Process A can also handle the findings of Markman (1973). The definition of a collection specifies its constituents, e.g., a set of parents and a set of children are the constituents of a family. A precedure CONSTITUENT could be inserted into LINK; it would call SETSEARCH separately for each constituent.

Processes A and B account well for the gross data. They could be tested more stringently by testing their processing sequences against detailed records.

References

Ahr, P. & Youniss, J. Reasons for failure on the class-inclusion problem. Child Development, 1970, 41, 131-143.

Carson, M. T. & Abrahamson, A. Some members are more equal than others: The effect of semantic typicality on class-inclusion performance. Child Development, 1976, 47, 1186-1190.

Isen, A. M., Riley, C. A., Tucker, T. & Trabasso, T. The facilitation of class inclusion by use of multiple comparisons and two-class perceptual displays. Paper presented at the meeting of SRCD, Denver, Colorado, April 1975.

Markman, E. The facilitation of part-whole comparisons by use of the collective noun "family." Child Development, 1973, 44, 837-840.

McGarrigle, J., Grieve, R. & Hughes, M. Interpreting inclusion: A contribution to the study of the child's cognitive and linguistic development. Journal of Experimental Child Psychology, 1978, 26, 528-550.

McLanahan, A. G. The class-inclusion problem: An information processing interpretation. Senior Honors Thesis, Princeton University, 1976.

Trabasso, T., Isen, A. M., Dolecki, P., McLanahan, A. G., Riley, C. A. & Tucker, T. How do children solve class-inclusion problems? In R. S. Siegler (Ed.), Children's Thinking: What Develops? Hillsdale, N. J.: Erlbaum, 1978.

Wilkinson, A. Counting strategies and semantic analysis as applied to class inclusion. Cognitive Psychology, 1976, 8, 64-85.

Winer, G. A. An analysis of verbal facilitation of class inclusion reasoning. Child Development, 1974, 45, 224-227.

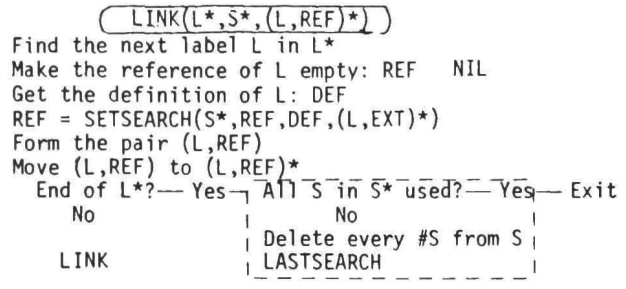


Figure 1. Modified flow chart of the LINK procedure used by both Process A and Process B (boxed section needed only for Process A).

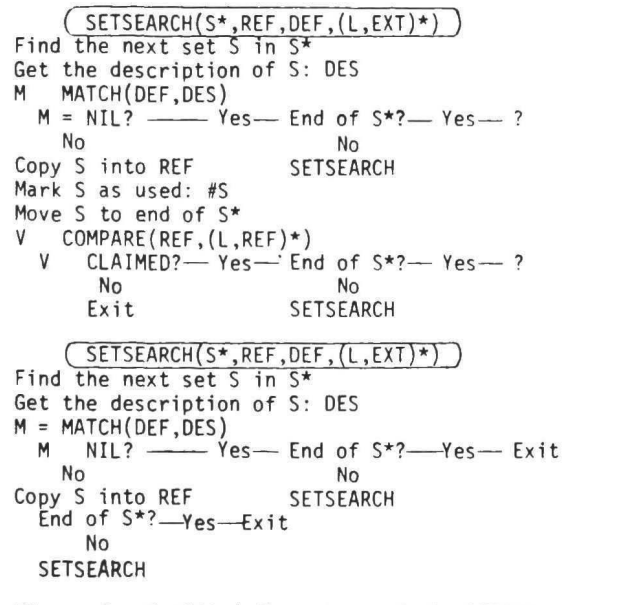


Figure 2. Modified flow chart of the SETSEARCH procedure used by Processes A (top) and B (bottom).

Standard task; Carson & Abrahamsen (1976) typical-typical condition. 5 horses & 3 dogs.
L*: "Horses, Animals" S*: HORSES, DOGS
L1: "Horses"/HORSES-Match-OK/ S*: DOGS, #HORSES
L2: "Animals"/DOGS-Match-OK/ S*: #HORSES, #DOGS
Result: 5 "horses" & 3 "animals" so "More horses"

Ahr & Youniss (1970); Winer (1974). 5 horses 3 dogs.
L*: "Horses, Dogs, Animals" S*: HORSES, DOGS
L1: "Horses"/HORSES-Match-OK/ S*: DOGS, #HORSES
L2: "Dogs"/DOGS-Match-OK/ S*: #HORSES, #DOGS
L3: "Animals"/HORSES-Match-But claimed, so try more /DOGS-Match-Ref of HORSES & DOGS is OK/
Result: 5 "horses" & 8 "animals" so "More animals"

McLanahan (1976). 4 horses, 2 dogs, 2 cats, 2 pigs.
L*: "Horses, Animals" S*: HORSES, DOGS, CATS, PIGS
L1: "Horses"/HORSES-Match-OK/
S*: DOGS, CATS, PIGS, #HORSES
L2: "Animals"/DOGS-Match-OK/
S*: CATS, PIGS, #HORSES, #DOGS
Have reached end of L*, but some sets unused.
Delete used sets to get S*:CATS,PIGS. Do LASTSEARCH.
S1: CATS/"Horses"-Mismatch/"Animals"-Match-OK/
S2: PIGS/"Horses"-Mismatch/"Animals"-Match-OK/
Result: 4 "horses" & 6 "animals" so "More animals"

Carson & Abrahamson (1976) atypical-atypical condition. 5 bees & 3 flies.
L*: "Bees, Animals" S*: BEES, FLIES
L1: "Bees"/BEES-Match-OK/ S*: FLIES, #BEES
L2: "Animals"/FLIES-Mismatch/#BEES-Mismatch/
Have reached end of L*, but one set unused.
Delete used set to get S*: FLIES. Do LASTSEARCH.
S1: FLIES/"Bees"-Mismatch/"Animals"-Mismatch/
Result: 5 "bees" & 0 "animals" so "More bees"

Carson & Abrahamson (1976) atypical-typical condition. 5 bees & 3 dogs.
L*: "Bees, Animals" S*: BEES, DOGS
L1: "Bees"/BEES-Match-OK/ S*: DOGS, #BEES
L2: "Animals"/DOGS-Match-OK/ S*: #BEES, #DOGS
Result: 5 "bees" & 3 "animals" so "More bees"

Carson & Abrahamson (1976) typical-atypical condition. 5 horses & 3 flies.
L*: "Horses, Animals" S*: HORSES, FLIES
L1: "Horses"/HORSES-Match-OK/ S*: FLIES, #HORSES
L2: "Animals"/FLIES-Mismatch/#HORSES-Match-But claimed, and no more sets to try. What to do?
Result: varies. Some compare the 5 HORSES to the 3 FLIES and answer "More horses"

Figure 3. Process A as a model of younger children's performance in several class inclusion studies: Highlights of possible sequences.

Standard task; Carson & Abrahamson (1976) typical-typical condition. 5 horses & 3 dogs.

L*: "Horses, Animals" S*: HORSES, DOGS
L1: "Horses"/HORSES-Match/DOGS-Mismatch/
L2: "Animals"/HORSES-Match/DOGS-Match/
Result: 5 "horses" & 8 "animals" so "More animals"

Ahr & Youniss (1970); Winer (1974). 5 horses 3 dogs

L*: "Horses, Dogs, Animals" S*: HORSES, DOGS
L1: "Horses"/HORSES-Match/DOGS-Mismatch/
L2: "Dogs"/HORSES-Mismatch/DOGS-Match/
L3: "Animals"/HORSES-Match/DOGS-Match/
Result: 5 "horses" & 8 "animals" so "More animals"

McLanahan (1976). 4 horses, 2 dogs, 2 cats, 2 pigs.

L*: "Horses,Animals" S*: HORSES, DOGS, CATS, PIGS
L1: "Horses"/HORSES-Match/DOGS-Mismatch/
CATS-Mismatch/PIGS-Mismatch/
L2: "Animals"/HORSES-Match/DOGS-Match/
CATS-Match/PIGS-Match/
Result: 4 "horses" & 8 "animals" so "More animals"

Carson & Abrahamson (1976) atypical-atypical condition. 5 bees & 3 flies.

L*: "Bees, Animals" S*: BEES, FLIES
L1: "Bees"/BEES-Match/FLIES-Mismatch/
L2: "Animals"/BEES-Mismatch/FLIES-Mismatch/
Result: 5 "bees" & 0 "animals" so "More bees"

Carson & Abrahamson (1976) atypical-typical condition. 5 bees & 3 dogs.

L*: "Bees, Animals" S*: BEES, DOGS
L1: "Bees"/BEES-Match/DOGS-Mismatch/
L2: "Animals"/BEES-Mismatch/DOGS-Match/
Result: 5 "bees" & 3 "animals" so "More bees"

Carson & Abrahamson (1976) typical-atypical condition. 5 horses & 3 flies.

L*: "Horses, Animals" S*: HORSES, FLIES
L1: "Horses"/HORSES-Match/FLIES-Mismatch/
L2: "Animals"/HORSES-Match/FLIES-Mismatch/
Result: 5 "horses" & 5 "animals" so "Same"

Figure 4. Process B as a model of older children's performance in several class inclusion studies: Highlights of possible sequences.

MOPs and Learning

Roger C. Schank
Yale University
Computer Science Department
PO Box 2158
New Haven, CT 06520

This paper is an attempt to sketch out some of what MOPs are about. It is taken from Schank (in press).

> A MOP is an orderer of scenes.
> A scene is a memory structure that groups together actions with a common goal, a common time, and some other common thread.
> It provides a sequence of very general actions. Specific memories are stored in scenes, indexed with respect to how they differ from the general action in the scene.

Scenes actually point to specific memories. MOPs do not. MOPs merely point to scenes. Scripts are particularly common instantiations of scenes. Thus, a scene consists of a generally-defined sequence of actions, while a script groups together particular realizations of the generalizations in a scene. Scripts package together particular realizations of scenes that have been known to frequently recur in a given context. Specific memories can be organized in terms of scripts also. This follows from the above, since a script is no more than a scene that has been colored (particularly instantiated) in a given way.

MOP's Defined

Since memories are to be found in scenes, a very important part of memory organization is our ability to travel from scene to scene. A MOP is an organizer of scenes. Finding the appropriate MOP, in memory search, enables one to answer the question 'What would come next?', where the answer is another scene. That is, MOPs provide information about how various scenes are connected to one another.

> A MOP consists of a set of scenes directed towards the achievement of a low level goal.
> A MOP always has one major scene that is the essence or purpose of the MOP.

There is a natural progression in terms of generality of structures that suggests itself:

> meta MOPs
> MOPs
> scenes
> scripts

Meta MOPs describe ordered progressions of scenes at their most abstract levels. As such they provide the stuff out of which MOPs are made. They do not actually contain memories. MOPs are less general descriptions of such progressions. The scenes they contain actually contain specific memories.

There are three kinds of scenes, physical, societal, and personal. Physical scenes represent a kind of "snapshot" of one's surroundings at a given time. Memories grouped in physical scenes provide information about what happened and how things looked.

Some MOPs refer to societal things rather than physical ones. M-CONTRACT is a MOP that organizes scenes that are not physically bounded. Thus, entities such as AGREE, or DELIVER, while behaving very much like scenes in a physical MOP, have no physical instantiation. They can happen anywhere and can take a great many different physical forms. Thus, a delivery of agreed upon services that fails to come about will be indexed under the DELIVER scene in M-CONTRACT. In this way, a failure of a department store to deliver a package that was paid for might remind one of a restaurant that required pre-payment and then failed to serve the desired food. Such reminding can only be accounted for by a memory organization that has scenes that are not exclusively physically bounded. DELIVER is an example of a societal scene, that is, one that may have many possible physical realizations.

Personal scenes are responsible for idiosyncratic behavior that is personally-defined. A personal scene is a scene whose common thread is a particular goal that belongs to the person whose scene it is. Any private plan to achieve ones own ends that is liable to repeat itself frequently is a possible personal scene.

This division in scenes is parroted by a similar division in MOPs. Physical MOPs can contain scenes that seem societal in nature, but what is actually happening is that one event is being governed by two scenes. Thus, for example, both M-CONTRACT which is a Societal MOP, and M-AIRPLANE, which is physical share a PAY scene. But each relates to different aspects of that event. In other words, "paying" can be seen as both a physical event and as a social event. Different MOPs provide expectations in each case. These expectations will coalesce to some degree, providing uniform expectations. Events confirming those expectations will be remembered in terms of both of the scenes that were active.

What is the difference between !$AIRPLANE (that is, our prior view of a script) and M-AIRPLANE (that is, our current view of a MOP)?

The Difference Between MOPs and Scripts

A MOP is an ordered set of scenes.
A script (1977 version) is an ordered set of scenes

BUT---
The definition of scene is different in each case.

For a MOP, a scene is a structure that can be shared by a great many other MOPs.
For a script (1977 version) a scene was particular to a given script and was not accessible without using that script.

A script (new version) is scene-specific. No script transcens the boundaries of a scene.

Now, to make this specific, let us actually look at M-AIRPLANE and !$AIRPLANE. Recall that !$AIRPLANE was more or less a list of an entire airplane trip. It included making the reservation, getting to the airport, checking in, riding in the plane, eating the meal and so on. In SAM and FRUMP these things were all stored in a complex structure, complete with optional tracks, under the name !$AIRPLANE.

But, now we wish to be able to make generalizations, get reminded across contexts and within contexts, and in general bring whatever relevant information from memory that we can find to help us in processing an input. To do this, we need structures that are far more general than a detailed complex list of events. For example, getting someplace by car, and making reservations by telephone are two scenes that were part of $AIRPLANE that could not possibly be part of M-AIRPLANE. The reason for this is that one could easily confuse one trip in a car to visit a friend who lives near the airport, with a trip to the airport that was intended to enable one to fly someplace. Similarly, one could easily confuse a phone conversation maiking airline reservations with one making hotel reseevervations. In fact they might well be the same conversation.

The problem with our old conception of scripts was that much too much that could have been defined generally, and that is likely to be stored in a general fashion in memory, was defined specifically as a part of a particular script. When one takes away everything that could have been defined generally from $AIRPLANE one is left with the things specific to $AIRPLANE, namely getting on the plane, being seated, being served a meal and so on. The above entities are the scripts that we now believe in. That is, M-AIRPLANE is a structure that, like any MOP, organizes a set of scenes. One of these scenes is SITTING IN THE PLANE (SITP). This scene has in it a number of scripts specific to that scene. These include $(SITP)EATING, $(SITP)MOVIE, and so on. Experiences that occurred within them, that is while those scripts were directing processing, that did not coincide with the expectations generated by that script, would be encoded as failures and indexed within that script.

M-AIRPLANE fills one strand of the meta MOP-TRIP. It consists of the following scenes:

M-AIRPLANE's scenes

CHECK-IN + WAITING AREA + BOARDING +
SIT-IN-THE-PLANE + DEPLANE +COLLECT-BAGS

Each of the scenes used by M-AIRPLANE is constructed as generally as possible. We should point out that it is people who are doing the construction of these scenes. One of the scenes of M-AIRPLANE is something called WAITING AREA. Now it is reasonable to ask, is this the same as the scene as WAITING ROOM in M-PROF-OFFICE-VISIT? Clearly such answers depend upon the experiences a memory has had and the decisions about what is like what (its generalizations) that it has made. It is perfectly plausible that a memory that had been to a doctor's and a lawyer's office and had constructed a scene WAITING ROOM, might upon its first encounter with an airport, see the waiting area as a version of WAITING ROOM. And, of course it might not.

Our point is that the possibility for such generalizations, for interpreting a new experience in terms of what it believes to be its most relevant old one, must exist for a memory. In order to do this, scenes must be memory structures in their own right, disassociated form the structures they are used with in processing. Thus MOPs as we have outlined them must be the kinds of memory structures we need. Scripts, in the old version of them, were too restrictive in this regard. This does not mean that scripts do not exist of course. Some of the experimental work on scripts relates to MOPs as we have now defined them and some of it relates to our new, more restricted definition of them.

Learning

Higher level learning and generalization takes place by indexing a given expectation failure in terms of the MOPs and scenes that were active at each of the three levels of analysis whenever the expectation failure occurred.

One key problem that a theory of memory must explain is what to do when an expectation fails. Consider again the Legal Seafood case (first discussed in Schank, 1980). After processing an episode at Legal Seafood, we would want to have detected a MOP-based expectation failure and have so indexed it. Why is this a MOP-based failure and how does a system know what structure to alter? The MOP M-RESTAURANT indicates the order of occurrance of scenes in a sequence. One way that a MOP can fail is by having the ordering of scenes that it predicts turn out to be wrong. In Legal Seafood, the PAYING scene comes immediately after the ordering scene. Thus M-RESTAURANT would be marked, at least initially, with an index after ORDER that PAY came next in this particular instance. But, just simply marking M-RESTAURANT is not enough.

The main question that is generated by any expectation failure is: What alteration of the structure that generated that expectation must be made? There are three possibilities, alteration, reorganization, and the construction of a new structure.

Consider our visitor to Burger King and MacDonald's. A first encounter with Burger King, for a person whose knowledge structures contain only the standard M-RESTAURANT, would produce an expectation failure in the order of ORDER, SEATING, PAY. When multiple failures occur, it is a good bet that it is because the MOP being used was of little value. Thus, in a situation of multiple failure, a new MOP must be constructed. This construction is complex since it involves reworking the existing MOP to create the new one. This is done by altering the MOP first, and the scenes second, as follows.

As in the Legal Seafood example, in Burger King PAY goes right after ORDER. In fact, we might expect a reminding here if the Legal Seafood episode came first. We have an additional problem with respect to M-RESTAURANT in that the SEATING scene follows PAY and ORDER. Further there are some script expectation failures too. For example, $RESTAURANT-ORDER is not usually done while standing.

The first thing that must be done then is to construct a new MOP. To construct a new MOP, we start with the scenes of the old MOP and reorder them according to the new episode. This is easy in the case of what we will temporarily call M-BURGER KING. The problem is that while the scenes may be the same, the scripts are different. A scene describes what takes place in general. And, in general, what takes place in a regular restaurant and a fast-food restaurant is the same. But the specifics are different. We do not want to use the scripts associated with M-RESTAURANT therefore. The problem then is to construct new scripts. Actually, this is hardly a problem at all. The new script is identical to the first Burger King episode. The real problem is to alter the scenes.

At this point we have a new MOP, M-BURGER KING, that contains the scenes ENTER + ORDER + PAY +SEATING with very specific scripts attached to each scene. Two problems remain. First we must encode the scripts correctly in the scene. Second we must generalize M-BURGER KING to the MOP that is more likely to be the one of greatest use, namely M-FAST FOOD. These two problems are related.

The scene alteration problem depends, after all, on how a scene is constructed in the first place. Let us then, by way of example, consider the ORDER scene.

ORDER as we have said, is a scene that is used by a great many MOPs. Some of these include: M-RESTAURANT, M-SHOPPING, M-PROVIDE-SERVICE, M-OFFICE, M-TELEPHONE-BUYING, M-TRAVEL-AGENT. The scene ORDER, in order to be used by this diverse set of MOPs, must be written in as general a way as possible. ORDER is one of those scenes that is both physical and societal. That is, it expresses both the generalizations that are valid when someone is physically ordering something, and those that pertain to the relationship between the participants in an ORDERing situation. Below, we have the physical scene ORDER. It looks a lot like a script, but without any particulars. Particular scripts, pointed to by ORDER, fill in the details (or 'color') the ORDER scene. Here then, is one possible view of ORDER:

```
Participants: actor, agent
Props: desired object or service;
   medium of MTRANSing
Preconditions:
  actor can MTRANS to agent
  agent can be assumed to have ability to
       get object or do service
  agent has willingness to get object or
       do service
Actions:
 actor establishes MTRANS linkage
 actor attracts attention of agent
 actor questions possibility of service being
          performed or object being delivered
 actor questions price of object or service
 actor states desire to agent
 agent agrees to comply
 agent tells actor when compliance will be complete
```

The role of a script attached to a scene is to color the scene with the particulars of that scene. In other words, a script is a copy of a scene with particulars filled in. For a script to be used, a copy of the scene is made that alters the scene in appropriate ways, leaving intact the parts of the scene that fit perfectly.

Knowledge that is acquired from restaurants about ordering that applies to all kinds of ordering will be known to so apply because the piece that was acquired will have been copied unchanged from the scene. The only way such knowledge can apply across the board to all ordering is if it relates to a non-restaurant specific portion of the script. In other words, expectation failures that are script-specific are stored in terms of the script itself. But, expectation failures that were due to expectations that were derived from a scene piece that was directly copied are stored in terms of the original scene.

To see how the scene-script relationship looks in practice lets consider the script !$RESTAURANT-ORDER. When !$RESTAURANT-ORDER colors ORDER it takes each line in it and either copy it directly or alters it to suit the script. For example, the precondition:

```
   agent has willingness to get object or
            do service
```

is a line in ORDER. !$RESTARANT-ORDER colors this line by adding the information that a waitress, can be, because it is her job, assumed to be willing.

Similarly the actions:

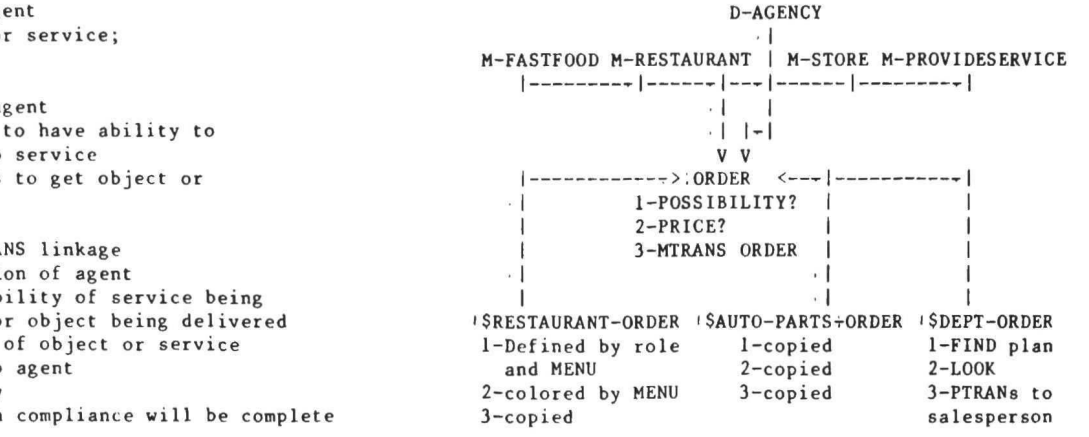
```
  actor questions possibility of service being
        performed or object being delivered

                    and

  actor questions price of object or service
```

are taken care of by the definition of a restaurant in the first case, and by a menu in the second. That is, !$RESTAURANT-ORDER colors ORDER by replacing an abstract line about price with information about reading a menu. Thus, these lines are altered in this script to reflect known information about restaurants.

Let's take a look at how a problem in ORDER can cause memory to be changed. Suppose we have a person who orders in a restaurant and finds that he isn't served because he spoke in an uncouth manner. Initially, this person might have a memory structure that looks like this:



This diagram shows how the scene ORDER is used by three scripts. Those scripts have copied the information in ORDER and either replaced it by coloring, or copied it directly. Three MOPs that use ORDER as one of its scenes are also shown. In addition, the generalization of the goal behind ORDER, namely D-AGENCY, is also connected to ORDER. This relationship is close to a hierarchical superset relationship. The MOP connection is one of filler to empty slot.

Now let's consider what happens when this person's nasty order goes unfilled in the restaurant. First, there is an expectation failure. The episode is indexed off of !$RESTAURANT-ORDER in slot 3. But, as slot 3 has been copied directly from the ORDER scene, this index is moved up to that level. When the same failure occurs in an auto parts store, or a fast food restaurant, or any other script that copies slot 3 directly from ORDER, a reminding occurs. This second instance causes a re-evaluation of the expectation that has failed. This re-evaluation causes an attempt to explain the failure.

In this case, the explanation is that servers don't like to be ordered nastily. Finding such explanations is an extremely complex process. Often they are not easily discoverable. We may need to be told. We may never find out. But, when we do find an explanation, it causes a local fix to be made that enables the person whose experience it was to modify ORDER in slot 3 accordingly. This

allows every MOP that uses ORDER to have that fix incorporated in it without doing a thing. The new, altered, ORDER is simply used by any MOP that previously used the old ORDER. In other words, this hypothetical person should now know to ask his wife to cook him things in a polite way and so on.

Now let's consider the Burger King example again. The problem in constructing M-BURGER KING is to take each scene that that MOP uses and treat each action that occurs within it in terms of its deviation from the baseline scene. Thus, $BURGER-KING-ORDER is built by noting how the actions observed in the first experience with Burger King differ from the ORDER scene.

The problem is, of course, that we want this MOP to be M-FAST FOOD. To get this MOP to be built, it is necessary to index M-BURGER KING in terms of M-RESTAURANT. The reason for this is as follows: Consider a patron entering MacDonalds. We want this patron to get reminded of Burger King. To put this another way, we want the patron to know to use M-BURGER-KING and not M-RESTAURANT. How can this be accomplished? One way is to index M-RESTAURANT at the point of its failed expectation relevant to Burger King, in this case noting that the scene ordering was different in a particular way. Thus, M-RESTAURANT must now have in it a marker recalling the past expectation failure and directing the processor where to go for help in further processing.

After this rerouting of processing has occurred a few times in the same way, the reminding ceases to occur. At that point M-BURGER KING has been transformed into a MOP with entry conditions of its own, that is, one that can be called in for use without even seeing it as a type of restaurant. To put this more generally, a new MOP is grown at the point where its conditions for use have been detected so that it can be called up independently from the MOP in which it originated as an expectation failure. Thus, after a few trials, M-RESTAURANT and M-FAST FOOD are independent MOPs.

In general then, expectation failures that are MOP-based, will initially just produce markers valuable for reminding. However, if the failure is radical enough, a new MOP must be constructed immediately.

REFERENCES

Schank, R. C. (1980) Language and Memory. COGNITIVE SCIENCE, Vol 4 no. 3, 243-284.

Schank, R. C. (In Press) DYNAMIC MEMORY: A Theory of Learning in Computers and People.

Retrieving General and Specific Information from Stored Knowledge of Specifics

James L. McClelland
University of California, San Diego

We often attribute the human ability to generalize from past experience to the use of stored representations (schemas, prototypes, etc.) in which generalizations are explicitly represented. This view is very appealing, but it raises two problems. First, there needs to be some mechanism for arriving at generalizations that are not stored explicitly, since it is unlikely that memory contains explicit representations that anticipate all of the possible generalizations we might ever wish to make. Second, we must explain how generalizations which are stored explicitly were obtained in the first place.

A mechanism that could induce generalizations from stored representations of specific objects or events could solve both these problems at once. It would explain how we could generalize when no explicit generalization is stored, and it would also suggest how we might have induced those generalizations which are stored.

Such a mechanism would also force us to consider whether we really store generalizations explicitly at all. If we can generate generalizations from stored representations of specific objects when we need to, explicit representation of these generalizations might turn out to be unnecessary.

Medin and Shaffer (1978) have suggested a first step toward the kind of mechanism I have in mind. Their model explains how we can assign a category label to a new object, based only on stored knowledge of the properties of previously encountered objects and the category labels that have been assigned to them. Their ideas can be extended to suggest how we may be able to do such things as answer questions about the general characteristics of classes of objects we have experienced before, and to fill in plausible default values for unspecified attributes of new exemplars.

The basic idea is that representations of previously-experienced exemplars stored in memory are activated via a spreading activation mechanism. Activated exemplars themselves activate representations of their properties. Mutually exclusive property values compete so that properties which are supported by a large subset of the active instances of the category are reinforced and become strongly active while those which are not are suppressed. Such a mechanism has recently been proposed by Glushko (1979) to account for our ability to construct apparently rule-guided pronunciations of nonwords (e.g., MAVE) without actually having any rules, and has been used by Rumelhart and me (McClelland and Rumelhart, in press; Rumelhart and McClelland, in press) to account for facilitation of perception of letters in words and nonwords. In both of these applications, the activation/competition mechanism is used to generate apparently rule-governed performance from stored knowledge of specific words.

I will illustrate the mechanism I am proposing by showing how it can be used to generalize from stored representations of specific objects. The representations of the objects are highly simplified, and are not sufficient to capture the varieties of structure of real objects. It is not my intention to advocate the representation. Rather, I use it to explicate the generalization mechanism, which is the main focus of interest here. We shall see that, even with a simplified representational system, the activation and competition mechanism can construct the general properties of classes of objects from stored knowledge of exemplars. It can also generalize along an indefinite number of different lines, retrieve the specific characteristics of particular exemplars, and fill in plausible default values for missing properties.

Table 1
The Jets and The Sharks

Name	Gang	Age	Edu	Mar	Occupation
Art	Jets	40's	J.H.	sing.	pusher
Al	Jets	30's	J.H.	mar.	burglar
Sam	Jets	20's	COL.	sing.	bookie
Clyde	Jets	40's	J.H.	sing.	bookie
Mike	Jets	30's	J.H.	sing.	bookie
Jim	Jets	20's	J.H.	div.	burglar
Greg	Jets	20's	H.S.	mar.	pusher
John	Jets	20's	J.H.	mar.	burglar
Doug	Jets	30's	H.S.	sing.	bookie
Lance	Jets	20's	J.H.	mar.	burglar
George	Jets	20's	J.H.	div.	burglar
Pete	Jets	20's	H.S.	sing.	bookie
Fred	Jets	20's	H.S.	sing.	pusher
Gene	Jets	20's	COL.	sing.	pusher
Ralph	Jets	30's	J.H.	sing.	pusher
Phil	Sharks	30's	COL.	mar.	pusher
Ike	Sharks	30's	J.H.	sing.	bookie
Nick	Sharks	30's	H.S.	sing.	pusher
Don	Sharks	30's	COL.	mar.	burglar
Ned	Sharks	30's	COL.	mar.	bookie
Karl	Sharks	40's	H.S.	mar.	bookie
Ken	Sharks	20's	H.S.	sing.	burglar
Earl	Sharks	40's	H.S.	mar.	burglar
Rick	Sharks	30's	H.S.	div.	burglar
Ol	Sharks	30's	COL.	mar.	pusher
Neal	Sharks	30's	H.S.	sing.	bookie
Dave	Sharks	30's	H.S.	div.	pusher

I will illustrate the features of the model by considering how it can be used to retrieve information about the members of two gangs called the Jets and the Sharks. Characteristics of hypothetical members of these two gangs are listed in the Table 1.

The model's knowledge of these individuals is captured in a node network. Each node is a simple processing device which accumulates excitatory and inhibitory inputs from other nodes continuously and adjusts its (real-valued) output to other nodes continuously in response, much as a neuron adjusts its rate of firing in response to a varying pattern of excitatory and inhibitory inputs.

The model has a node for each of the individuals it knows and a node for each of the properties or attributes these individuals may have. The former are called instance nodes and the latter are called property nodes. There is a property node for each individual's name, one for each gang, one for each age range, one for each educational level, and so on. Property nodes are arranged into groups or cohorts of mutually exclusive values. The instance nodes are also treated as a cohort of mutually exclusive nodes. In the following Figure, the instance nodes have been placed in the center with the property nodes all around. Nodes within a cohort (bounded region) are mutually inhibitory.

The system's knowledge of an individual consists simply of an instance node and a set of bi-directional excitatory links between it and the nodes for the properties that individual is known to have. For example, the system's representation of Lance is an instance node with mutual excitatory connections to the name node "Lance", the gang membership node "Jet", the age node "20's", the education node "Junior High", the marital status node "married", and the occupation node "burglar".

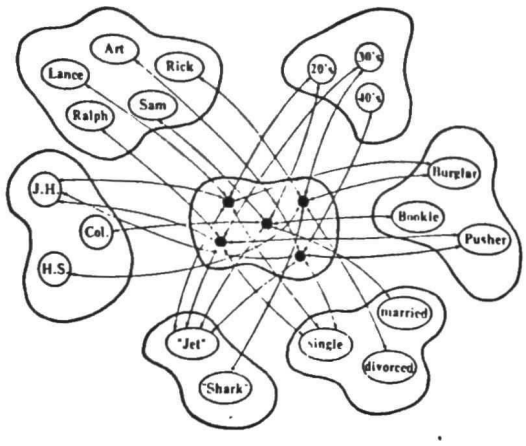


Figure 1. The representation of several of the individuals listed in Table 1.

Filling in Properties in Response to a Probe

The system is queried by presenting it with a probe. For example, to find out about the properties of the Jets we can probe the system by activating the property node "Member of Jets". A probe might be a name or any other single property, or it may consist of a list of properties.

Before a probe is presented, each node is assumed to be at rest, with an activation value below 0. Probe presentation causes an excitatory input to be applied to each node specified in the probe. This excitatory input is allowed to stay on, and as time passes it drives the activations of the specified property nodes above 0, into what is called the active range. Active nodes send excitatory signals to the instance nodes they are linked to and send inhibitory signals to the other nodes in the same cohort. These signals are graded, and their strength is proportional to the source node's activation. As processing continues, some of the instance nodes become active. They then begin to excite the property nodes they are connected to, and to inhibit all the other instance nodes. Eventually, property nodes not present in the probe may become activated.

The excitation and inhibition processes are allowed to go to equilibrium. At this point, the system has generally activated property nodes for properties not specified in the probe. These activations are the system's response to the probe. If all of the active instance nodes "agree" on a property, the node for that property will tend to be strongly activated. On the other hand, if they all specify different values within the same cohort, many values will become partially activated and they will all tend to cancel each other out. In any case, what is filled in can then be used as a basis for overt response to the probe. For example, a statement of the typical age of the members of the Jets could be based on the resulting pattern of activation over the age nodes. I will go through some examples of what the system fills in in response to various probes after giving a few more details of the working of the model.

Quantitative Details

The net input to node i at time t is given by:

$$input_i(t) = p_i(t) + E\sum_j e_{ij}(t) - I\sum_j i_{ij}(t).$$

$p_i(t)$ stands for the probe input to node i. It is set to +.2 if the probe drives node i and to 0 otherwise. The $e_{ij}(t)$ are the activations of the active excitors of node i and the $i_{ij}(t)$ are the activations of the active inhibitors of node i. The constants E and I are simply weights which modulate the excitatory and inhibitory effects of the input. Their values (.05 and .03) are the same for all nodes except as noted below.

The effect of the net input to node i is modulated by the current activation ($a_i(t)$). If the net input is excitatory (i.e., greater than or equal to 0) then the effect is

$$effect_i(t) = (M-a_i(t))input_i(t)$$

If the net input is inhibitory (i.e., less than 0) then the effect is

$$effect_i(t) = (a_i(t)-m)input_i(t)$$

Here M stands for the maximum possible activation of the node and m stands for the minimum. This formulation ensures that the activation of each node stays between the maximum and minimum values, which are set to 1.0 and -.2 respectively.

There is a tendency for the activation of each node to decay at some rate D back to its resting value R. This tendency is subtracted from the effect of the net input to the node to determine the rate of its activation:

$$d(a_i(t))/dt = effect_i(t) - D(a_i(t)-R) .$$

The values of D and R are .05 and .1.

Simulation

The behavior of the system described above is simulated on a digital computer by using discrete rather than continuous time. One every tick of the discrete clock, the activations of each node are adjusted to reflect the effects of the activations of other nodes at the end of the previous tick. The time slices are kept thin by using small values for E, I, and D, so that the approximation to a continuous system is quite close.

Examples of the Model's Behavior

Let us examine the system's response to the probe "Member of the Jets". Presentation of the probe causes the "Jet" node to become active, and this in turn sends activation to the instance nodes of all of the members of the Jets. As they become active they send excitation to the nodes for their properties. These nodes in turn reinforce the activations of those jets with active properties. After about 200 cycles the pattern of activation over the property nodes has stabilized at the following values:

Name:	--			
Gang:	Jets	.869		
Age:	20's	.663		
Ed:	J.H.	.663		
Mar:	Sing.	.663		
Occ:	Pusher	.334	Bookie	.334
	Burglar	.334		

Activations for instance nodes are omitted to save space. All property nodes not mentioned are below zero activation. Based on these activations the model could generate a list of its conception of the typical properties of the Jets. In the case where only one possibility is active, the system would simply report that value. Where multiple possibilities are active, it could either list the set of possibilities or make a probabilistic choice from among the alternatives.

In this case the active age, education, and marital status properties are the ones which are typical of the Jets. Though no Jet has all three of these properties, 9 out of 15 of the Jets are in their 20's, 9 have only Junior High educations, and 9 are single. The occupations are divided evenly among the three possibilities. Thus, the model tends to activate the node on each dimension which is most typical of the members of the gang, even though it has never encountered a single instance with all of these properties, and has no explicit representation that the Jets tend to have these properties.

An interesting feature of the model is that it can retrieve the typical properties of any subset of individuals matching an arbitrary conjunction of specifiable properties. For example, we can probe with the properties "Age in 20's" and "Junior High Education". Four individuals have these two properties. All of them are Jets and Burglars by trade. Two of them are married and two divorced. The response of the system reflects these facts:

Name:	Lance	.127	John	.127
	Jim	.094	George	.094
Gang:	Jets	.732		
Age:	20's	.855		
Ed:	J.H.	.862		
Mar:	Mar.	.589	Div.	.389
Occ:	Burglar	.721		

In this case the instance nodes for the four individuals matching the probe become strongly enough activated to drive the activations of the corresponding name nodes above threshold. Lance and John get more active than Jim and George because the instance node for Al, a married individual who is very similar to Lance et al., becomes slightly activated, thereby boosting the activation of the "married" node and causing Lance and John to gain a slight edge.

The model can also be used to retrieve the properties of a particular individual. In so doing, it exhibits the tendency to fill in "default" values for unknown properties of an instance. To illustrate this, we can delete the link between the instance node for Lance and the "burglar" node and then see what happens when we present the name "Lance" as a probe. The Lance name node becomes active and excites the corresponding instance node. This excites the nodes for the known properties of Lance. These then excite the nodes for other individuals who share these properties. Finally, they in turn excite the nodes for properties that they share. When the pattern of activity finally stabilizes (in about 400 cycles) the model has filled in an occupation for Lance.

Name:	Lance	.799		
Gang:	Jets	.710		
Age:	20's	.667		
Ed:	J.H.	.704		
Mar:	Mar.	.552	Div.	.347
Occ:	Burglar	.641		

The value filled in is shared by the other individuals who are most similar to Lance (namely John, Jim and George). At equilibrium the different marital situations of these individuals are also reflected in the pattern of activation. The model has blended its representation of Lance with its representation of other very similar instances.

This kind of blending can be a good or a bad thing, of course. It is sometimes important to know what we really know about something rather than what we might plausibly assume based on our knowledge of similar things. Fortunately, a single parameter of the model -- the strength of mutual competition among instances -- determines whether the model will tend to fill in values from partial activations of related instances. If active instances inhibit each other strongly, then the most strongly activated instance will tend to dominate the pattern of activation and keep other instances from "contaminating" the information retrieved. In the Lance example, if the strength of instance-to-instance inhibition is increased from .03 to .05, the instance node for Lance dominates the instance nodes and the others are kept from getting active so they cannot activate the missing occupation or the competing marital status. Thus, the model can either retrieve what is actually known about an instance or it can fill in missing properties from the common properties of similar instances.

In summary, the model I have described is capable of generalizing along a number of different lines about the shared properties of specified subsets of familiar objects. It can also retrieve what it knows about specific instances, and, if desired, fill in plausible default values for unknown properties of the retrieved individuals. It can induce generalizations as it needs them across novel partitions of the knowledge base. Since these are many of the behaviors which have led workers in various fields of cognitive science to assume we explicitly store generalizations, the model raises the possibility that this assumption, however plausible, may not necessarily be true in all cases.

There are many more steps to be taken, of course. For one thing, the model needs a representational system which can capture more highly structured knowledge. How the model can be extended in this way while preserving its interesting properties is currently being explored.

Acknowledgements

The work reported here was supported by NSF Grant BNS79-24062. I would like to thank Steve Draper, George Mandler, and Don Norman for very useful comments.

References

Glushko, R. J. The Organization and activation of orthographic knowledge in reading words aloud. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 674-691.

McClelland, J. L., & Rumelhart, D. E. An interactive activation model of the effect of context in perception, part I. Chip Report 91, Center for Human Information Processing, University of California, San Diego. La Jolla, California, 1980.

Medin, D. L. & Shaffer, M. M. Context theory of classification learning. Psychological Review, 1978, 85, 207-235.

Rumelhart, D. E., & McClelland, J. L. An interactive activation model of the effect of context in perception, part II. Chip Report 95, Center for Human Information Processing, University of California, San Diego. La Jolla, California, 1980.

CAN *IF* BE FORMALLY REPRESENTED?

D.S. Brée

Any attempt to arrive at a formal semantics for natural language must at least provide a mapping of the function words into the chosen formal representation. One of the function words that has given considerable problems to logicians is *if*. Even today there is a debate concerning whether or not *if* is equivalent to material implication, $\supset$, in the propositional logic; see for instance the journal Analysis. It has long been recognised that equating *if* with $\supset$ leads to difficulties. I will look at two proposals to cope with this within the confines of traditional propositional logic: one is an older logician's approach, Reichenbach's (1947) connective interpretation; the other is a recent proposal for a 'natural' logic put forward by the psychologist Braine (1978). As neither is satisfactory I will next look at two approaches involving modal logic: the well known 'strict implication' of Lewis and Langford (1932) and an improvement due to Stalnaker (1968), both of which are also unsatisfactory. This will bring me to a consideration of the 'possible worlds' logic which appears to be particularly suitable for conditional propositions. Regrettably, it once again turns out that a second proposal by Stalnaker (1975), namely that *if* p,q may reasonably be inferred from $\tilde{p}$ *or* q and vice-versa, is not proven. At this point one is tempted to abandon the attempt to capture the natural use of *if* in a formalism and to agree with Grice's highly convincing notion that the difference between *if* and material implication can be accounted for by certain conversational implicatures to which all discourse is bound. However, I find that a rather simple interpretation of *if* in the first order predicate calculus, supplemented with a convention to differentiate asserted from presupposed propositions, appears to meet all the standard objections.

The well known problems that arise by equating *if* p,q with $p \supset q$ are:

a. *Affirmation through denial of the antecedent* is permissable for $\supset$, i.e.
$\tilde{p} \rightarrow p \supset q$, but not for *if*, e.g.
A: If God exists we are free to do what we want.
B: How do you come to that conclusion?
A: By knowing that God doesn't exist.

b. *Affirmation through assertion of the consequent* is permissable for $\supset$, i.e.
A: If the government reforms there will be rioting in the streets.
B: How do you know?
A: Because there are always street riots here at this time of year.

c. *Denying* material implication permits only one true state, i.e.
$\sim(p \supset q) \rightarrow p \,\&\, \tilde{q}$, but denying a conditional may be something else, e.g.
A: It's not so that if God exists we are free to do what we want.
B: Are you claiming that God exists and we aren't free to do what we want?

Similar problems arise when *or* is equated with $\vee$, a situation which led Reichenbach (1947) to propose that a 'connective' interpretation be given to $\vee$ in order to make it equivalent to *or*. This interpretation requires that all possibilities must remain open. Applying this idea to $\supset$ would require that it must not be possible to do away with any of the three residual statements for the truth of $p \supset q$, i.e. $(p\&q)$, $(\tilde{p}\&q)$ and $(\tilde{p}\&\tilde{q})$, in order for $\supset$ to be equated with *if*. Now both denying the antecedent, $\tilde{p}$, and asserting the consequent, q, rule out two of these three and so no *if* statement may be used. Moreover, denying an *if* p,q may be denying that all three residual statements are open. So Reichenbach's proposal copes with the three well known problems. However, it is open to criticism as it fails to distinguish the relative importance of the three residual statements for an *if* statement:

d. *Homogeneity:* $\tilde{p}\&q \rightarrow p \supset q$ but not *if* p,q as seen in an example of Conditional Perfection (Geis & Zwicky, 1971)
A: If you mow the lawn, I'll give you 5 dollars.
B: (Returning after a minute) May I have the 5 dollars now?
A: But you can't have mowed the lawn already!
B: I haven't, but you said there was a possibility of my getting the 5 dollars anyway ($\tilde{p}$ and q).

It is not only logicians who have tried to adapt propositional logic so that $\supset$ can be equated with *if*, psychologists also are interested. For example let us consider a recent attempt by Braine (1978) to set up a 'natural' logic based on 18 'natural schemata' which he claims are available for reasoning. These include for example modus ponens but not modus tollens, which is thus not available as a [illegible] but only follows from several steps. Braine equates *if*

with /, which is equivalent to an inference in his system. But it can be shown that / suffers from all three of the traditional problems of ⊃, see Figure 1. Note in particular the ease with which the truth of the consequent q, can be used to derive p/q. I believe that we should conclude from both the logicians' and psychologists' attempts to adapt propositional logic so that *if* can be equated with ⊃ that this is not the way to go.

Figure 1

Proofs that Braine's system of Natural Logic suffers from the traditional problems

Affirmation from $\tilde{p}$				Affirmation from q				Denial to $p \& \tilde{q}$			
##	from	by	gives	##	from	by	gives	##	from	by	gives
A1		N16	$\tilde{p}$	B1:		N16	$p \supset q$	C1		N16	P(p/q)
A2	1	N 5	F(p&q)	B2:	1	N 2	q				
A3	1,2		$\tilde{p}$/F(p&$\tilde{q}$)	B3:	1,2		(p&q)/q	C2		N16	q
								C3	2	B 6	p/q
								C4	1,3	N 1	(p/q)&F(p/q)
A4		N16	p	B4:		N16	q	C5	2,4		q/((p/q)&F(p/
A5		N14	F(p&$\tilde{q}$)	B5	3,4	N18	p/q	C6	5	N 7	$\tilde{q}$
A6	4,5	N12	F(q)	B6:	4,5		q/(p/q)				
A7	6	N 8	q					C7		N16	$\tilde{p}$
A8	4,5,7		(p&F(p&$\tilde{q}$))/q					C8	7	A11	p/q
								C9		N 1	(p/q)&F(p/q)
A10		N16	F(p&$\tilde{q}$)					C10	7,9		$\tilde{p}$/((p/q)&F(p/
A11	9,10	N18	p/q					C11	10	N 7	p
A12	10,11		F(p&$\tilde{q}$)/(p/q)								
								C12	5,11	N 1	p&$\tilde{q}$
A13	1,12	N17	p/(p/q)					C13	1,12		((p/q)/(p&q)

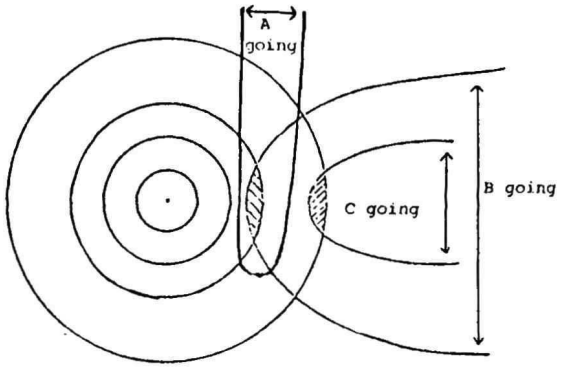
N1 to N20 are Braine's natural schemas, most of which are obvious: N16 is 'Assuming'.

Attempts to equate *if* with a modal operator have also met with difficulties. For instance the well known proposal that *if* might be equivalent to strict implication put forward by Lewis and Langford (1932) partially overcomes all four of the problems of material implication. Strict implication is defined as asserting the impossibility of $p\&q$, i.e. $\sim\Diamond(p\tilde{q})$ Strict implication is thus a stronger concept than material implication; the latter can be deduced from the former. However, it still suffers from stronger versions of the two problems of affirmation as $\sim\Diamond p \rightarrow \sim\Diamond(p\tilde{q})$ and $\Box q \rightarrow \tilde{}\Diamond\tilde{q} \rightarrow \tilde{}\Diamond(p\tilde{q})$ but it is not so that from either the impossibility of the antecedent or from the necessity of the consequent that a conditional *if* statement may be deduced, as can be seen by appropriately modifying the examples given for the affirmation problems above. At one point Stalnaker (1968) attempted to improve on the Lewis and Langford proposal by equating *if* p,q with $p>q$; his corner, >, entailed material implication and is entailed by strict implication, i.e. $p>q \rightarrow p \supset q$ and $\sim\Diamond(p\tilde{q}) \rightarrow \Box(p \supset q) \rightarrow p>q$. Unfortunately this leaves it open to the same criticism as strict implication, for instance the impossibility of the antecedent is sufficient for the corner, i.e. $\sim\Diamond p \rightarrow \Box(p \supset q) \rightarrow p>q$. Modal logic has not offered us a formal representation of *if*, although it has come closer than simple propositional logic.

It is then with high expectations that we turn to the logic of 'possible worlds' which seems designed to cope with conditionals. In brief this postulates a universe of possible worlds arranged in an ordered series of sets such that each set contains all the worlds of the previous set and some more worlds besides. The initial set contains one world, which is usually interpreted as our world as it is. Worlds added in going from one set, K, to the next set, K+1, are further removed from our world than all worlds in set K. These sets are the contexts in which sentences are interpreted. For a conditional sentence the appropriate context is the first set which contains a world in which the proposition underlying the antecedent is true. A nice consequence of this is that *if* is no longer a transitive relation as can be seen by examining Figure 2. Even 'if B goes to the party, A will go' and 'if C goes, B will go' it does not follow that 'if C goes, A will go'; A may well want to go to meet B but not in the unlikely circumstances of C's going as well.

Figure 2

Possible worlds representation of if, p, q is not transitive



Suppose: 'If B's going to the party, A's going'
But not: 'If C's going to the party, A's going'

Figure 3

A formal version of Stalnaker's proof

To Proove: 'P̃ *or* Q therefore *if* P, Q' is a reasonable inference

Proof: Suppose 'P̃ *or* Q' is appropriate in K.
S2: Suppose 'P̃ *or* Q' is accepted in K.
S3: So P&Q is appropriate in K.
S4: So P is appropriate in K.
S5: If X is appropriate and Y is accepted in K, '*if* X, Y' is accepted in K.
S6: So '*if* P, then P̃ *or* Q' is accepted in K.
S7: So '*if* P, then Q' is accepted in K.
S8: So 'P̃ *or* Q therefore *if* P, Q' is a reasonable inference.

It is within this formalism that Stalnaker (1975) set out to show that, although *if p,q* may not be truth functionally equivalent to $\tilde{p}$ *or q* the one is a reasonable inference from the other and vice-versa. A reasonable inference occurs when for all contexts, K, in which the premise is appropriate and acceptable the conclusion is also. A proposition, P, is appropriate in K if there exists at least one world in K in which P holds; P is acceptable if it holds in all worlds in K. A formal version of Stalnaker's proof that '~P or Q therefore if P,Q' is a reasonable inference is reproduced in Figure 3. Unfortunately it is flawed as it assumes what it is setting out to prove, namely that *if* P,Q is equivalent to P⊃Q in step 5. The result is not surprising as even the 'impossible worlds' formalism of *if* accepts any appropriate conditional in which the consequent is necessarily true, the second of the traditional problems of affirmation.

Those of you who never believed in the logical basis of natural language will by now be thinking 'I told you so' and those convinced in the formal program will be busy finding a new formalism. Perhaps we should follow a middle road, for example the path marked out by Grice (1967) in his William James lectures. He virtually divided the problem into two parts: retain a simple formal representation for natural language connectives, including *if*, and account for deviations between the formalism and the normal usage by postulating a set of Indirect Conditions, IC. Then *if* ≡ ⊃ + IC. The IC's are the cancellable part of the meaning of *if*, e.g. part of the IC is that the speaker doesn't know the truth values of either the antecedent nor the consequent but these are cancellable as in

> I know where Smith is and what he is doing; all I'll say is that if he's in London he's attending the meeting.

Although they are cancellable they are not detachable, i.e. it isn't possible to find another formulation which is equivalent to *if*-IC, e.g. both

> Either Smith isn't in London or he's attending the meeting.
>
> It isn't the case that Smith's in London and not attending the meeting.

contain the same IC. Such IC's Grice called Conversational Implicatures, CI's, which hold for conversation in general not only for *if*. His first CI is the maxim of quantity: a speaker does not say less than he knows. Thus a speaker will not use *if p,q* when he knows $\tilde{p}$ or *q* for certain, which avoids the two affirmation problems. His third CI is more specific: *if p,q* has an implicature 'supposing *p*, then *q*'. This CI is also present with $\tilde{p}$ *or q*. It enables the problems of denial and homogeneity to be avoided. However, it doesn't have the generality that one would like from a CI. Would it be too uncharitable to say that it ducks the issue?

With a certain trepidation I would like to propose a formalism for *if p,q* in the first order predicate calculus

F1: $\forall w(P(w) \supset Q(w)) \;\&\; [\exists(x,y)(P(x)\&\sim Q(y))]$

in which the proposition P is the proposition derived from *p* in the context K, and P(w) is true if P is the case in world w. The universal, ∀, and existential, ∃, operators operate over all worlds in the context set K, following the 'possible worlds' formalism. The necessity for differentiating *p* and P is not only that the proposition underlying a sentence depends on the context in which that sentence is interpreted, but also because the protasis and apodosis of a conditional are not necessarily equivalent to sentences, they sometimes cannot stand alone, e.g.

> If anyone has a malignant cancer of the backbone, they'll be dead within 6 months.
>
> If Alexander was afraid, I didn't notice it.

As Ryle (1950) pointed out *if* sentences contain statement indents, not statements. They can be used for making inferences, but are not in themselves inferences.

This formalism avoids the three traditional problems of equating *if* with ⊃. Let W be our world. Then neither ~P(W) nor Q(W) is sufficient to affirm *if p,q* unless W is the only world in K. Nor can it be confirmed by $\tilde{p}$ nor by *q* as these two extremes are ruled out by the 'presupposition' of *if*, given in F1 between square brackets. The problem of denial is also avoided, as denying *if* denies the assertion $\forall w(P(w)\supset Q(w))$ which is equivalent to $\exists w(P(w) \;\&\; \sim Q(w))$, but the world that satisfies this denial is not necessarily our world. However, this formalism fails to avoid the homogeneity problem: there is no difference between a world for which PQ is true and one for which $\tilde{P}Q$ is true. This problem can be avoided if we are prepared to use strict implication rather than material implication. We then define *if p,q* to be represented by

F2: $\forall w(P(w)\rightarrow Q(w)) \;\&\; [\exists(x,y)(P(x) \;\&\; \sim Q(y))]$

in which the same conventions hold as above, and → is strict implication. Specifically strict implication has a known truth value only when the antecedent is true. Thus P→Q is true for PQ and false for $P\tilde{Q}$, which avoids the homogeneity problem. It remains to be seen if there are other problems which this formalism is not capable of handling.

At first glance this formulation for *if* seems to cope with problems that arise for 'other' interpretations of *if* than the standard. Most notorious of these is the counterfactual in which the premise is claimed to be false of the actual world, but from a false proposition anything follows! As Lewis (1973) has shown the 'possible worlds' approach, which I have here adapted, can cope with counterfactuals. There is no claim in F2 that P(W) is false (or true),

e.g.

> If Paul had stuck to his plan, he'd (still) have been famous.

(Examples are adapted from the Brown corpus of American English, Kučera and Francis, 1967). F2 also copes with factuals in which both P(W) and Q(W) hold, e.g.

> If Wilhelm Reich is the Moses who has led them out of the Egypt of sexual slavery, Dylan Thomas is the poet who offers them the Dionysian dealectic of justification for their indulgence in liquor.

As P(W) is true Q(W) must hold. But contrary to the possible worlds formulation, the context, K, must not be confined to W as then $\neg\exists w(\neg Q(w))$. Nor is there any problem with Austin's (1961) stipulative use of *if*, e.g.

> There are some biscuits on the table, if you want some.

as again the truth status of Q(W) is not necessarily open but may be true. And I think it will handle cases of doubtful presupposition as in

> It made him conspicuous to the enemy, if it was the enemy.

Here the interpretation of *q* is problematic unless P is true, but since F2 uses strict implication this does not matter.

I do not claim that the use of F2 for *if* can decide which of these 'interpretations' is actually the case. Rather that decision should not rest upon *if*, but must be made using other aspects of the sentence, e.g. what the listener knows that the speaker knows. This is particularly the case for the use of *if* within the scope of performative verb, e.g.

> He promised vengence on V.L. if ever the chance came his way.

What I do say is that F2 *permits* these different interpretations, with the exception of the performative.

REFERENCES

AUSTIN, J.L. (1961)
Ifs and cans. In J.O. Urmson and G.J. Warnock, eds., *Philosophical papers of J.L. Austin*. London; Oxford University Press.

BRAINE, M.D.S. (1979)
On the relation between the natural logic of reasoning and standard logic. *Psychological Review, 85*, 1-21.

GEISS, M.L. and ZWICKY, A.M. (1971)
On invited inferences. *Linguistic Inquiry, 2*, 561-566.

GRICE, H.P. (1967)
William James Lecturers, Harvard University. Published in part as "Logic and conversation", P. Cole and J.L. Morgan, eds., *Syntax and Semantics, Vol. 3: Speech Acts*. New York; seminar Press, 1975, 41-58.

KUCERA, H. and FRANCIS, W. N. (1967)
Computational analysis of present-day American English. Providence; Brown University Press.

LEWIS, C.I. and LANGFORD, C.H. (1932 and 1959)
Symbolic Logic. 2nd ed. New York; Dover.

LEWIS, D.K. (1973)
Counterfactuals. Cambridge, Mass.; Harvard University Press.

REICHENBACH, H. (1947)
Elements of symbolic Logic. London; Macmillan.

RYLE, G. (1950)
'If', 'so' and 'because'. In M. Black, ed., *Philosophical analysis*. New York; Cornell University Press.

STALNAKER, R.C. (1968)
A theory of conditionals. In: N. Rescher, ed. *Studies in logical Theory*. Oxford; Blackwell. Reprinted in E. Sosa, ed., *Causation and counterfactuals*. Oxford University Press, 1975.

STALNAKER, R.C. (1975)
Indicative conditionals. *Philosophia, 5*, 269-286.

TO SEE .AND. NOT TO SEE, THAT IS THE QUESTION

E. Andreewsky[1], A. Andreewsky[2]
G. Deloche[1] and D. Bourcier[3].

[1]INSERM U 84, Hopital de la Salpetriere, 47, Bd de l'Hopital, 75014, PARIS-FRANCE.

[2]LIMSI-CNRS, Orsay FRANCE

[3]ERA 430 CNRS-FRANCE.

Both psychological and computational theories of the lexicon usually consist of the specifications of the semantic relations between words. The present paper proposes an alternative, and simpler lexicon which does not include the particular meanings for each item in it. This lexicon consists instead of pointers to a general knowledge data-base, on the basis of which the meanings of words are to be computed.

World knowledge is indeed required to understand properly any word. This is demonstrated by examples like the following : in "a green salad" vs."a green cadaver", the color of "green" is not the same! But the requirement may be more obvious for a lexical ambiguity. For instance, to account for the meanings of the verb "study" in : "Francois studies english" vs. "Chomsky studies english", one cannot help but take into account much knowledge about how english is not the french people (as Francois refers to) mother tongue, on one hand, and, on the other hand, about Chomsky's scientific activity. The psychological meaning of words does indeed depends on world knowledge : it is more likely to result from some processing (taking this stored knowledge into account) than to be just retrieved out of any structure.

In this paper, we will, first, give a computational model for the resolution of lexical ambiguity; it is grounded on an abstract "lexicon", with pointers to the knowledge data-base. We will then present psycholinguistic data which call for such a preprocessing as the above abstract lexicon may provide.

RESOLUTION OF POLYSEMIES. AN A.I. APPROACH.

Given a (french) sentence, including a lexical ambiguity, m, we want to disambiguate m. The present resolution of this problem is based on an automatic information retrieval system, in natural language, so called "SPIRIT" (A. Andreewsky et al, 1980)which is currently use in Paris, with various data-base.

The "world knowledge" of SPIRIT is its stored documents (the french texts D_1, D_2... D_n).

The "lexicon" of SPIRIT is very poor : it contains mainly morphological and syntactical properties of words.

SPIRIT computes the "distances" d_i, between any given french sentence $\underline{s}$ (that is, any request to the system) and its n stored documents; these distances d_i enable the system to answer to $\underline{s}$, by means of an hierarchically arranged list of numbers :

$$A_s = (i_1, i_2, \ldots i_m)$$

i.e. a set of weighted numbers pointing to the documents Di_1, Di_2, ... Di_m, answering the best to the request $\underline{s}$.

SPIRIT includes a syntactical disambiguator: For instance, a word such as "can" (he can open the can) has two different entries in its lexicon, whereas a word such as "bachelor" has only one. To handle the p different meanings of such polysemies, the following steps are taken :

- Usual dictionnary definitions of words (one definition for a regular word, p for a p-polysemous word) are input, as given "request" to SPIRIT; the system's answers (one to p for each word) are sets of numbers A.; these sets are linked to each lexical entry, and stored. The lexicon of the system include now, together with morphological and syntactical data, pointers to the knowledge data-base the A.

- If a polysemous word m has the p meanings : m_1, m_2.. m_p, its lexical entry include the sets : A_1, A_2.. A_p. Given, now, the word m in a sentence $\underline{s}$ the following operations enable its disambiguation, in a very simple way :

- A_s is the answer of SPIRIT to the sentence $\underline{s}$.
- A_1, A_2.. A_p are the answers of SPIRIT to the p different definitions of the meaning of m.

the greatest of the following intersections : $A_1 \wedge A_s$, $A_2 \wedge A_s$ ..... $A_p \wedge A_s$ give which definition of m shares a maximum of related documents with the sentence $\underline{s}$. It is the meaning of m in the given sentence $\underline{s}$.

In order to exemplify how the system actually works, consider the disambiguation of the word : "instruction". In french, this word has three meanings :

1) instruction = "teaching or education, the taking in charge of school-age children."

Given the first definition, the answer of the system is :

A_1 = 120, 513, 519, 1829, 611, 1361, 207. 42\1\2

2) instruction = "Proceedings which bring a case or a law-suit to trial".

to this definition, the answer of the system is :

A_2 1761, 1760, 1376, 1393, 1369, 1723, 276.

3) instruction = "Directions, instructions or informations given for indicative purposes".

answer of the system :

A_3 1802, 2144, 1761, 1367, 2720, 2490.

(all the numbers here refers to a data-base of 3000 laws, chosen for the present experiment).

Now, given the following sentence :

s "Qui se charge de l'instruction de dossiers de recouvrement d'impots" (Who is in charge of the examination of records in tax collection cases), including "instruction", the answer of the system is:

A_s = 276, 1367, 2121, 1761, 1760, 1376, 1393, 1361.

What is the meaning of "instruction" in the sentence $\underline{s}$? The intersection of the answer of the system to $\underline{s}$, with the answers to the three definitions of "instruction" are :

$A_1 \wedge A_s$ 1361
$A_2 \wedge A_s$ 276, 1761, 1760, 1376, 1393
$A_3 \wedge A_s$ 1367, 1761.

Thus, the 2nd meaning of "instruction" triggers the greatest intersection, therefore it is the meaning of this word in the context at hand. An example of the use of "instruction" in the same meaning is taken out of the document N° 1761 . "le service d'assiette procede à l'instruction de la demande"(the tax-assessment service proceeds to investigate claims)

Here, there is a lexicon without any semantic information which enables, however (out of its pointers to the knowledge data-base) to handle lexical ambiguities in sentences.

* *
*

PSYCHOLINGUISTIC EXPERIMENTS.

How people process lexical ambiguities? In general, there is two contreversed models to deal with this problem (cf Carey P.W. et al, 1970). requiring a lexicon from which meanings are retrieved. If more than one meaning are retrieved, the wrong ones are then inhibited; the inhibition hypothesis is also invoked to explain subluminary experiments were people are able to implicitely use some semantic properties of words they are unable to report, or even to notice (cf D.A. Allport, 1977)

We will present two psycholinguistic experiments which require either a meaning-retrieval and inhibition mechanism, or, alternatively, some abstract lexicon such as the one above.

Experiment 1.

Written words were tachistoscopically presented in pairs, with visual masking conditions, in a speed which allowed subjects to report at most one of them.

All the words were nouns, and they include the french homograph : "fils" (which means either "son" or "threads"); this word is not an homophone, and when it means "son" it is uttered /fis/, and /fil/, for "threads". The written word "fils" was displayed together with either the word "father" or the word "needle". (E. Andreewsky et al, 1978). The subject's sole pronounciation /fis/ or /fil/ testifies in favor of the implicit resolution of the written polysemy; this resolution was found to be in accordance with the "ungrasped" co-displayed noun (father or needle)

The subjects' utterances testifies therefore in favor of implicit use of the "meaning" of words, displayed but not understood.

Experiment 2.

M.L. Albert et al (1973) propose a method so-called the "odd-word-out-test". Subjects are asked to point to the odd item in a written list, such as "hat" in ; "cat, dog, pig, hat, wolf". There are alexic subjects, who can neither read aloud nor understand (i.e. match with proper pictured items) any of the written nouns in the list, but can, nevertheless, perform the "odd-word-out-test". We have reported (G.Deloche et al, on press), in the context of the above task, the behavior of one of those alexic subjects : he gives 8 correct answers out of 10 lists of 5 written items, all of which he cannot understand. Therefore, since semantic cues are obviously required for the selection of odd items, such a patient displays, here again, the ability to make implicit use of the "meaning" of words he "sees", without being able to understand.

* *
*

The question arises then : how people can manage "to see and not to see" written items, that is make an implicit use of the meaning of written words which are not understood? This is the case in the two above experiments, such as in many others (Cf A. Marcel, on press).

Under a structural model of the human lexicon, from which meaning of words should be stored and retrieved, it is impossible to explain these experiments without a strongly ad hoc hypothesis on an inhibition mechanism, following the retrieval of a lexical meaning (explaining how one can make use of a word meaning, and not understand this word).

An alternative explanation can be carried out, grounded on such an abstract lexicon as described in the A.I. part of this paper. In the system described above, the lexicon does not include semantic informations, but only sets of pointers to the knowledge data-base. The retrieval of pointers is a preprocessing towards word understanding. It is clear that these pointers provide enough information -even not semantic- to explain how people may "see and not see" the meaning of words. For instance, in the case of the alexic patient, if such pointers would be retrieved, without further processing of the meaning of words, the patient will perfectly well be able to point to the odd word, out of an intersection between the pointers of the list, without having any idea on any written word meaning.

Therefore, our A.I. approach not only allows a very simple way to handle lexical ambiguities, but also provides a theory to explain how people can manage to see and not to see a word, that is to implicitely use semantic properties of non-understood written items.

REFERENCES.

- Albert, M.L., Yamadori, A., Gardner, H. and Howes, (1973), Comprehension in Alexia, "brain", Vol. 96, part 2, 317-328.
- Allport, D.A. (1977) On knowing the meaning of words we are unable to report : the effect of visual masking. In "Attention and Performance", Dornic (ed) Lawrence Erlbaum, New Jersey, 505-533.

Andreewsky, A., Debili, F., Fluhr, C. (1980). Apprentissage Syntaxe semantique lexicale."Revue du Palais de la Decouverte", Vol. 9, N° 83, 17-40.

Andreewsky, E., Deloche, G., and Kossanyi, P. (1979). Towards a procedural understanding of reading INS Conf., Noordwijkerhout, Holland. (in "Les processus de la Lecture", special issue, 1978).
- Carey, P.W., Mehler, J. and Bever, T.G. (1970) Judging the veracity of ambiguous sentences. "Journal of Verbal learning and Verbal Behavior", 9, 243-254.
- Deloche, G., Andreewsky, E., and Desi, M. (on press) Surface dyslexia : a case report and some theoretical implications to reading models. "Brain and Language".
- Johnson-Laird, P.N. (1980). Mental models in cognitive science. "Cognitive Science", 4, 71-115.
- Marcel, A.J. (on press). Conscious and unconscious reading : The effects of visual masking on word perception. "Cognitive Psychology"

E/MOTIONAL MEMORY AS A MEDIATING CONSTRUCT IN THE STUDY OF PERSON/ENVIRONMENT INTERACTION. Marisa Zavalloni, Department of Psychology, University of Montreal.

Recent trends in social and environmental psychology underlie the importance of studying psychological processes through which a person and his/her environment interact (Bronfenbrenner, 1979). At the same time, it has become more and more apparent that psychology has established a methodological and epistemological tradition which is not adapted to the study of interactive phenomena (Cronbach, 1975; Proshansky, 1970). The traditional approach of scientific psychology is based on the comparison of average responses measured on an aggregate, and it aims to elicit general properties of the psychological system, such as attitudes, values, and needs, which allegedly represent the basic parameters of hypothetical constructs through which it is hoped that one day it will be possible to explain the complex functioning of the mind. However, as Feldman and Lewontin (1975) note, there is a vast loss of information in going from a complex machine to a few descriptive parameters and an immense indeterminacy in trying to infer the structure of the machine from those few descriptive parameters. Moreover, the data thus obtained refer to average characteristics of a group and not to properties expressing some transaction between the person and his/her environment. The development of a social/ecological and interactive psychology will require a methodological perspective capable of producing a concrete research program having as prerequisite a much clearer understanding of what it means to study interactive phenomena as they operate in the real world. If the epistemological orientation in psychology today appears to be acquiring an interactive and constructivist outlook, it would be wrong to assume that we witness a harmonious development in this direction. On the contrary, as discussed elsewhere (Zavalloni and Louis-Guérin, 1979), we are in a period of transition and uneasiness, characterized by a widening gap between the theoretical reflection and its concrete applications.

The constructivist, interactive perspective in psychology has found persuasive defenders in recent years, yet very much has to be done to translate the implications of this perspective into a research model and to devise concrete methods for its application. The major problem is how to have access to the processes at work in the internal construction of reality of which environmental perception is one of the elements. The fundamental question to be raised in this connection is, are there empirically detectable structures responsible for these processes, or will this level of psychological functioning in the real world remain beyond the reach of psychologists?

Our research goal was to discover through the Social Identity Inquirer (SII) the relation between the objective social identity of a person and his/her internal operant environment (Zavalloni and Louis-Guérin, 1981). The SII represents a complex procedure based on a method called representational contextualization. The procedure comprises different phases each using different technics: free association, focused introspection and associative network analysis. The results obtained indicate that the linguistic encoding or categorization through which we evaluate the socio-cultural environment constitutes only one aspect of what is activated in the brain when a person enunciates these categories. The method of representational contextualization has permitted to identify several psychological components which accompany tacitly the evaluation categories people use to describe the environment such as images, collective and biographical memories and events that act as recoding features of the general stimuli used to produce the categories. These contextual components may be seen as constituting the psychological meaning of a category in contrast to its semantic meaning as provided by a dictionary or common sense. At the same time, they disclose the content of the internal operant environment.

The method of representational contextualization is designed to elicit the latent connections between a concept or category and images, thoughts, and experiences which constitute its background. The principal feature of this method is to provide a display of cognitive material which is never accessible to consciousness or research in its totality, but which is experienced in a segmented way at different times under particular situations of elicitation. The resulting combination of words, images, and memories whose connections have been found stable over time has been defined as conceptual-e/motional cluster and appears to constitute the units of representational thinking. What are activated are the ideational and experiential features which summarize the salient aspects of the transaction between the individual and his/her environment, and which in the natural situation inhabit the periphery of consciousness.

The conceptual-e/motional cluster as a unit activated when a respondent is asked to evaluate an element of the sociophysical environment includes the following components: 1) Representational unit (RU): category of concept, 2) Implicit operant referent, 3) Prototype images 4) Episodes (exemplifying memory), 5) Subjective meaning of RU.

These components can be described as follow:
1) The representational unit (RU) represents one of the categories or concepts used by a person to evaluate an element of the sociophysical environment (groups of identity, of alterity and feature of the physical environment. The sum of all the RUs constitutes the semantic repertory of a person, descriptive of his/her internal environment.
2) The implicit operant referent. The linguistic categories (RUs) used to describe a group in general or other environmental features are mediated by a subgroup and/or individuals important in the life space of the respondent. These were found, through repeated elicitations over time to be stable features of a respondent associative thinking. This result indicates the existence of a recoding mechanism by which objective groups are translated into particular subgroups and/or individuals evolving in a socio-physical environment. The respondent is rarely aware that his/her linguistic encoding is generated by these implicit operant referents and not by the manifest stimulus to which he/she responds. The sum of the implicit operant referents constitutes the sociophysical microcosm of a person.
3) The prototype images are also stable features of memory content and/or symbolic images associated to each RU as descriptors of the environment.

The sum of these prototype images is part of the microcosm of a person as well as an indicator of the motivational and ideological system of a person.

4) Episodes (exemplifying memory) are salient features of the group history and of the biography of a respondent. They represent those aspects of the socio-physical environment that are embedded in the personal and collective biography.

5) The subjective semantic refers to subjective connotation of the categories used (RUs), when applied to a particular referent. This subjective meaning emerges as a complex interaction between the general meaning of the linguistic categories and of the referent as a social object.

It should be emphasized that if the elements of the conceptual-e/motional cluster inhabit the periphery of consciousness, in the natural situation, this is due not to their intrinsic nature, but to the situation of elicitation. Under a different conditions of stimulation these background data may emerge as manifest response. This will occur, for instance, if a person is asked about preferred environments, some important events in his/her life, ideological preferences, etc.

But this direct elicitation will not permit us to see the processes through which this personal and subjective memory content operates in the evaluation of the sociophysical environment, at a given moment on time.

What is specific to the representational contextualization method is that it elicits the unconscious links which exist between a rational discourse, an ideological stance, and the affect, images which are associated to a personal biography. The identification of the conceptual e/motional cluster thus as a unit of analysis, constitutes a step toward the empirical study of the relation between language and thinking in the area of self, alter and society.

A structural analysis described elsewhere (Zavalloni and Louis-Guérin, 1981) permits us to detect patterned relations between conceptual-e/motional clusters, leading to the identification of what could be called the intrapersonal ideational system. This dynamic structure, constituted by a limited number of subsystems called sociomotivational nuclei, represents the articulation between one's aspirations and desires and one's evaluation of the sociophysical environment. The sociomotivational nuclei seem to represent some invariant modality for defining, responding to, and adapting to the sociophysical environment.

This constitutes an affective-cognitive Gestalt which could be defined as e/motional memory. Since Tulving (1972) there has been a tendency to distinguish between two types of long-term memory, episodic and semantic. Episodic memory contains stored information concerning episodes and events situated in time. Semantic memory refers to concepts and abstractions. E/motional memory appears to contain those aspects of episodic and semantic memories which are invested with affect (possessing motivational properties and orienting a person's action in the real world). The concept of e/motional memory as emerged to describe a dynamic structure empirically linking our memory of the world, our existential project, and our evaluation of the sociophysical environment. What we have identified as the internal operant environment can be considered the content of e/motional memory.

It is important to note that the psychological material thus obtained does not reflect some properties of the organism as it is usually understood and measured through traditional methods (attitude scales, interviews, etc.). Here the organism is seen as indissociable from its umwelt or internal environment, defined as those elements of the sociophysical environment which are enmeshed in its motivational system. They are a mixture of stability and change as the individual proceeds through life. The internal operant environment (the content of e/motional memory) as a dynamic structure, stored in long-term memory, can provide a baseline for investigating the modalities of person/environment interaction in a particular research situation and as such could be considered as a mediating construct.

The uncovering of the conceptual-e/motional cluster as a unit of representational thinking, has permitted us to set as a goal the exploration of invariances and patterns where there appeared to be a continuously changing and elusive flow of consciousness, a domain which has been considered outside the reach of science, to be left to writers and poets. (Marsh, 1977)

References

Bronfenbrenner, U., 1979. The Ecology of Human Development. Cambridge: Harvard University Press.

Cronbach, L.J., 1975. Beyond the two disciplines of scientific psychology. Amer. Psychol., 30, 116-127.

Feldman, M.W., and Lewontin, R.C., 1975. The Heriditability Hang-up. Science, 190, no. 4220, 1163-1168.

Marsh, 1977. A framework for describing subjectivity of consciousness. In Zinberg, N., ed., Alternate States of Consciousness, New York: Free Press.

Proshansky, H.M. et al., 1970. The influence of the Physical Environment of Behavior: Some Basic Assumptions. In Proshansky, H.M., Ittelson, W.H. and Rivlin, L.G., Environmental Psychology: Man and his Physical Setting, New York: Holt, Rinehart and Winston, Inc.

Tulving, E., 1972. Episodic and Semantic Memory. In Tulving, E. and Donaldson, W., Organization of Memory, New York: Academic Press.

Zavalloni, M., 1975. Social Identity and the recoding of reality: its relevance for cross-cultural psychology. Inter. Jour. of Psychology, 10, no. 3, 197-217.

Zavalloni, M. and Louis-Guérin, C., 1978. Environment, perception et mémoire. Int. Rev. Appl. Psychol., 27, no. 1, 1-8.

Zavalloni, M. and Louis-Guérin, C., 1979. Social Psychology at the Crossroads: its encounter with cognitive and ecological psychology and the interactive perspective. Eur. Jour. of Social Psychology, 9, no. 3, 302-322.

Zavalloni, M. and Louis-Guérin, C., 1981. Identité et pensée sociale. (Submitted for publication).

THE PERCEPTION OF DISORIENTED COMPLEX OBJECTS

Steven P. Shwartz
Cognitive Science Program
Yale University
New Haven, CT 06520

Abstract

Subjects were presented with pictures of real-world objects at varying orientations and required to name them. Naming latencies increased with the angular departure of the pictured object from the object's canonical orientation. The results suggest that, at least when orientation-invariant features that uniquely identify the object are not present, subjects mentally rotate an internal representation of the object into its canonical orientation in order to complete the recognition process.

Introduction

Theories of visual perception are divided on the issue of how the human visual system recognizes objects at non-standard orientations. Some theorists argue that the visual system searches for orientation-invariant features and bases perceptual recognition on these features (e.g. Milner, 1974). The fact that certain common objects seem to be immediately recognizable at non-standard orientations is compatible with this view. Other theorists (e.g. Rock, 1973) argue that in order to recognize a disoriented object, the object must be first imagined in its canonical orientation. The fact that some complex objects (e.g. faces) are difficult to recognize when disoriented supports this position.

The classic experiments of Cooper and Shepard (1975) suggest that in order to imagine disoriented objects at their standard orientations subjects perform a mental rotation of an internal representation of the object. The Cooper and Shepard paradigm required subjects to judge whether or not a presented letter had a normal or mirror-reversed form. They found that the time to make this decision was a monotonic function of the angular distance of the orientation of the presented letter from the canonical orientation of the letter. This suggests that before it was possible to make the required judgement, it was first necessary to perform a mental rotation of an internal representation of the presented letter in order to imagine it in its canonical orientation.

However, as Cooper and Shepard noted, these results may have been due to a special strategy developed by subjects for the purpose of performing the normal vs. mirror-image judgement because normal and mirror-image letters can not be distinguished on the basis of distinctive features. For this reason, the Cooper and Shepard experiments are not strong evidence that mental rotation is normally used in the recognition of disoriented objects. In fact, Corballis, Zbrodoff, Shetzer, and Butler (1978) have shown that under certain circumstances (i.e. when a subject is required to determine whether a presented stimulus was a specified target letter or one of 5 distractor letters), the time to identify a letter is independent of the orientation of the presented letter.

Thus, it is clear that people are capable of performing mental rotations on internal representations of perceptual stimuli in order to normalize the orientation of the perceptual stimulus prior to recognition. However, it is equally clear that mental rotation is not always required for the recognition of disoriented forms. Under certain conditions, people are capable of utilizing distinctive features of perceptual stimuli in order to recognize disoriented perceptual stimuli. In the Cooper and Shepard study, subjects were forced to use mental rotation because of a lack of available distinctive features. In the Corballis et. al. study, it was possible that subjects were able to extract a small set of features that could be used for identification of the target letter and that these features were recognizable at all orientations.

The purpose of the present study is to determine the extent to which mental rotation is used in the perceptual recognition of disoriented objects that a person might encounter in the physical world. In the present study, subjects are presented with drawings of physical world objects at varying orientations and required to identify them by name. If mental rotation is used to identify disoriented stimuli, then the time required to identity each stimulus object should be a monotonic function of the angular distance of the presented stimulus object from the canonical orientation of the object. Subjects were tested on novel stimuli -- i.e. subjects had no foreknowledge of the indentities of the objects they would see -- in order to ensure that they could not use a strategy of searching only for a few particular features.

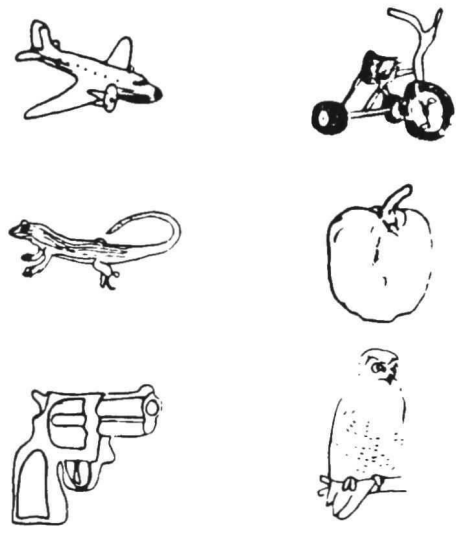


Figure 1. Examples of the stimulus objects used in the experiment.

Method

Subjects.

Twelve subjects from The Johns Hopkins University served as subjects in the experiment in partial fulfillment of a course requirement.

Materials and Apparatus.

The stimuli used in the experiment consisted of 48 xeroxed drawings of real-world objects, examples of which are shown in Figure 1. The stimulus set included 15 animals, 30 inanimate objects, and 3 well-known faces (Washington, Carter, and Reagan). Six versions of each of the 48 drawings

were constructed, one at each of the following angular departures (in degrees) from the object's normal upright: 0, 60, 120, 180, 240, and 300. The stimuli were presented to subjects individually in a two-field tachistoscope, subtending an average horizontal and vertical angle of 3.2 degrees.

Procedure.

On each trial of the experiment, the experimenter simultaneously said the word "READY" and pressed a button initiating the trial. After a .5 sec warning interval, the stimulus was presented and remained visible until the subject verbally named the depicted object. The subject's verbal response triggered a voice activated relay that stopped a response timer and the subject's response latency was recorded. Before the start of the experiment, the experimenter explained the operation of the voice key and gave each subject practice at triggering the voice key only at the time of response. Each subject was instructed to verbally name each stimulus as rapidly as possible without making errors. In order to reduce the ambiguity of possible stimulus names, subjects were instructed that the response given should be the name that one would normally use in referring to the stimulus object and, as an example were told that the correct response to a picture of an animal is the name of the animal (and not "animal").

Design.

Each subject received 48 test trials, one for each of the 48 objects. Each subject was tested on 8 trials at each of the 6 orientations. Thus, each subject was tested on only one of the 6 orientations for each stimulus in order to ensure that a novel stimulus was presented on each trial. For each set of 6 subjects, each of the 48 objects were tested at each of the 6 orientations. Each subject received 12 practice trials with stimulus objects that were not included in the set of test stimulus objects.

Results.

Responses were considered correct when the name given was appropriate and was not the name of a superordinate category. The largest number of errors made by a subject was 4 and error trials were not included in the analyses.

The latency data were analyzed separately with both subjects and objects as random variables. The direction of disorientation -- clockwise (60 and 120 degrees) vs. counter-clockwise (300 and 240 degrees) -- had no effect on response latencies [$F < 1$ for both subjects and objects], nor was there a significant direction x orientation (60 vs. 120 degrees) interaction [$F < 1$ for both subjects and objects]. For this reason, the data for the 60 and 300 degree orientations as well as for the 120 and 240 degree orientations were not distinguished in the analyses.

The primary results of the experiment are shown in Figure 2. As can be in Figure 2, there is a monotonic effect of the angular departure from the canonical orientation of the object on response latencies [$\min F' = 3.519$, $p < .05$]. This analysis was also run without inclusion of the 3 faces and the pattern of results did not change as a result of this analysis. This result is compatible with the view that normalization of orientation occurs prior to recognition.

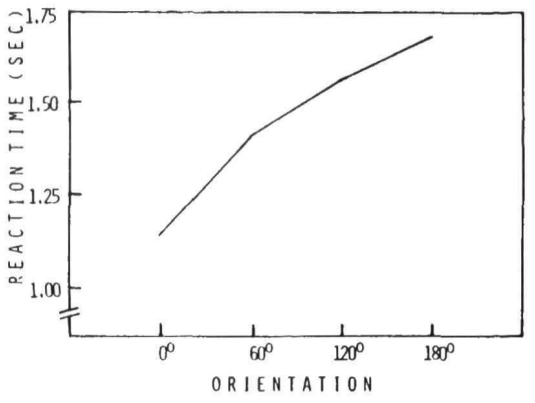


Figure 2. Reaction time as a function of the angle of departure from the canonical orientation of the object.

For each of the 48 stimulus objects, the best-fitting straight line relating the angular orientation of the object to response latency was computed using linear regression. The correlation coefficient of the regression provides a measure of how well the data are fitted by a linear function. The median correlation coefficient was .602, with 36 positive correlation coefficients and 12 negative coefficients. These data, along with the linear trend in evidence in Figure 2, constitute evidence that recognition difficulty increases as the degree of angular departure from the object's canonical orientation increases.

Discussion.

These results are consistent with the view that people use mental rotation to normalize their internal representations of visual stimuli in order to facilitate recognition. However, while a monotonic function of orientation was found for most of the stimuli used in the present experiment, there were stimulus objects for which such a trend was not found. Since orientation is manipulated between-subjects -- i.e. no subject sees a stimulus at more than a single orientation -- the non-monotonically-increasing trends could have been due to noisy data.

However, it is equally likely that for some of these stimuli, there do exist orientation-invariant features that are used to identify the object. For example, three objects for which a linear trend was not found were the owl, the alligator, and the gun. It is easy to imagine how the big eyes of the owl, the long teeth and curvy tail of the alligator, and the barrel and trigger of the gun can serve to uniquely identify each without mental rotation. These features are easy to spot at arbitrary orientations and are unique to these objects.

Corballis et. al. (1978) were similarly able to obtain identification functions for letters that were independent of orientation when the sets of positive and negative features were sufficiently restricted (by restricting the sizes of the sets of positive and negative letters). It is hypothesized that in the Corballis et. al. experiment these restrictions had the effect of increasing the cue validity of certain features of the letters that could be identified at arbitrary orientations to the point where these features could uniquely identify the letter. That is, Corballis et. al. were able to create a context in which the available orientation-invariant features were sufficient to uniquely identify the letter.

The results of the present experiment suggest the following algorithm for the identification of disoriented perceptual stimuli: When orientation-invariant features are available and are sufficient to uniquely identify an object, they are used. However, when such features are not available, or when these features do not uniquely identify the object, a mental rotation of the internal representation of the object is performed in order to extract features that are not orientation-invariant.

The author wishes to thank Pierre Jolicoeur for a helpful discussion of some of the issues involved in this research, Nancy Tang for collecting the data, and Ray Gibbs for comments on the manuscript.

References

Cooper, L. A. Shepard,R.N., Chronometric studies of rotation of mental images, in W.G. Chase (Ed.), Visual Information Processing, New York: Academic, 1973.

Corballis, M. C., Zbrodoff, N. J., Shetzer, L. I. & Butler, P. B. Decisions about identity and orientation of rotated letters and digits, Memory and Cognition, 1974, 6(2), 98 107.

Milner, P. A. A model for visual shape recognition, Psychological Review, 1974, 81, 521 - 535.

Rock, I. Orientation and form. New York: Academic Press, 1973.

The Role of Intrinsic Axes in Shape Recognition

Marianne Wiser
Massachusetts Institute of Technology

This paper presents a model for the mental representation of visual shapes which accounts for their recognition in different orientations, including novel ones. It is motivated by the following considerations.

Some shapes are easier than others to recognize after rotation in space. For example, it is easy to see that Fig. 1a and Fig. 1b represent the same shape in two different positions whereas Fig. 2a and Fig. 2b seem to represent two different shapes. This suggests that a rotation-invariant representation is more readily available for shape 1 than for shape 2. Shapes that are easily recognized in different positions are also perceived as in a specific orientation with respect to an external frame of reference (here, the page); for example, shape 1 looks vertical in Fig. 1a and oblique in Fig. 1b. Shape 2, on the other hand, is not perceived as in any specific orientation.

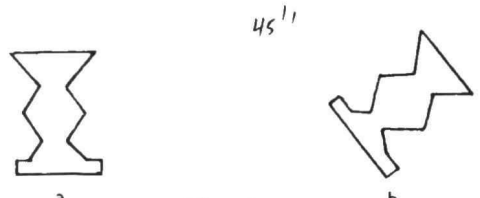


Fig. 1

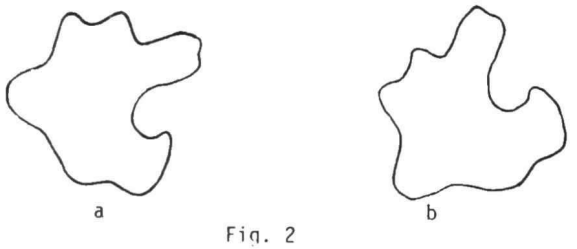


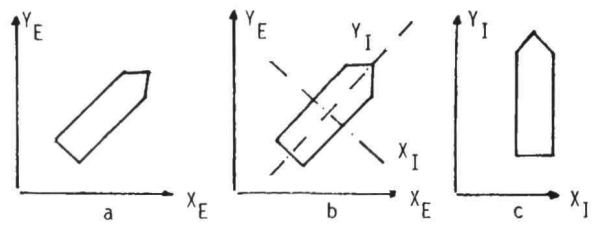
Fig. 2

Shapes like shape 1 are perceived as oriented because they have an axis of their own, determined by their geometrical properties (symmetry, elongation, parallel sides). The position of their axis relative to the side of the page determines their perceived orientation. I will call this axis "intrinsic" because it is fixed within the shape and exists independently of any other direction in space. In contrast, axes like the retinal or gravitational vertical, or the side of the page are "external" because their direction is independent of the shape being perceived. Shape 2 is not perceived as oriented because it lacks axis-determining properties.

Rock (1973) proposes a theory that explains why shapes like shape 2 are difficult to recognize after rotation. When a shape is perceived, it is described in a specific spatial frame of reference (e.g. the page); shapes are compared and recognized on the basis of this description. If the shape is rotated with respect to the frame of reference being used, its description changes and consequently, it is hard to recognize. Rock uses the notion of description within a frame of reference to explain why the change in retinal orientation that occurs when an observer tilts his head does not affect the percept of the shape, whereas the same change, when it results from tilting the shape itself (with the observer's head upright), generally makes the shape look different (as in Fig. 2). He suggests that the description tends to be performed in a gravitational, rather than retinal, frame of reference. As long as the shape has a fixed position in space, its gravitational description does not change and it will, therefore, look the same although its position on the observer's retina changes.

This theory explains why the world does not tilt when we tilt our heads, but it also predicts that we should not be able to recognize a shape rotated in space since, in that case, the description within both the gravitational and the retinal frames of reference change. A fortiori, it cannot explain the differences between shape 1 and shape 2. I propose that, when a shape has an intrinsic axis, it is used as a frame of reference to compute a description of the shape, the intrinsic description, which is independent of its position in the external frame of reference.

More precisely, the intrinsic axis and a perpendicular to it form a system of coordinates or frame of reference: the intrinsic frame of reference (Fig. 3). During the encoding of the shape,



(a) Description in terms of external coordinates (X_E, Y_E)
(b) Position of the intrinsic axis Y_I in the external frame
(c) Description in terms of intrinsic coordinates (X_I, Y_I)

Fig. 3

the shape is described in terms of intrinsic coordinates, i.e., its elements are localized within the intrinsic frame of reference. Such a description is rotation-invariant because the intrinsic frame of reference is fixed within the shape.

The intrinsic axis therefore plays a dual role in the perception of the shape. On the one hand, it indicates the position of the shape in the external frame of reference; on the other, it is the frame of reference in which an invariant description of the shape can be built. The intrinsic axis makes it possible to keep separate the invariant part of the information contained in an external description (the identity of the shape) from the source of variation in the external description (the orientation of the shape in the external frame).

The experiment reported here provides evidence that the intrinsic axis and the intrinsic description are part of the mental representation of shapes. In this experiment, I will not attempt to distinguish between retinal and gravitational frames. Both are external, as opposed to the intrinsic frame. The subjects in the experiment were tested with head upright so that the retinal and gravitational frames coincided. I will refer to this frame as retinal, to simplify.

Experiment

A learning and recognition paradigm was used to test the hypothesis that the recognition of shapes with intrinsic axes is based on their intrinsic descriptions. Two models for the memory representation of shapes are contrasted: the retinal encoding model and the intrinsic encoding model. In the retinal encoding model, the intrinsic axis plays no role in the processing of the shapes, which are treated like shapes without intrinsic axes. They are stored in memory as retinal descriptions (i.e.

spatial descriptions in terms of retinal coordinates). In the intrinsic encoding model, it is the intrinsic description which is stored in memory. Shapes with intrinsic axes were presented for learning in either a vertical, oblique, or horizontal position. They were simple two dimensional non-sense shapes (Fig. 4).

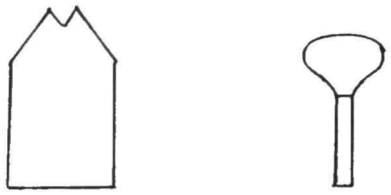


Fig. 4

45/4 The task was to recognize them among distractors when they were presented again in either the same or one of the other two positions. The subjects had to say "yes" as fast as possible if they recognized a shape they had learned and "no" if they did not. They were warned before the recognition session (but not before the learning session) that some shapes would be rotated. A "yes" shape was seen in one of 9 combinations of orientations: VV (vertical during learning/vertical during testing), VO (vertical during learning/oblique during testing), VH (vertical during learning/horizontal during testing), OV (oblique during learning/vertical during testing), etc.

For shapes seen in an oblique or horizontal position, the intrinsic frame of reference does not coincide with the retinal frame, and, thus, the intrinsic description differs from the retinal one. Depending on whether the retinal or the intrinsic description is used to represent the shape in memory, a different pattern of reaction times was predicted in each of the 9 conditions.

The retinal encoding model predicts that RT's should be faster when the learning and testing orientations are the same (i.e. when the retinal description of the target matches the description in memory) than when they differ. This is justified by numerous experiments in which shapes without intrinsic axes were used (Dearborn, 1899; Shinar and Owen, 1973; Rock, 1973). Therefore conditions VV, OO and HH should be equally fast and faster than conditions VO, VH, OV, OH, HV, and HO.

The intrinsic model on the other hand predict that shapes tested vertically should be faster to recognize than shapes tested in an oblique or horizontal position, irrespective of the orientation in which they were learned. This prediction is based on the assumption that the encoding of the intrinsic description requires more processing when the shape is oblique or horizontal than when it is vertical. The geometrical properties that determin the intrinsic axis, such as symmetry, are detected faster about a vertical than an oblique axis (Julez 1971) and the building of the intrinsic description itself involves a shift in perceptual frame of refe ence when the shape is not vertical. Therefore con ditions VV, OV and HV should be equally fast and faster than conditions VO, VH, OO, OH and HH.

Since the models tested in this experiment co cern the memory representation of the shapes, the analysis of the "yes" responses only will be considered. It was run on RT's in msec. for correct responses. Fig 5 shows that the pattern of RT's supports the intrinsic encoding model for horizontal as well as oblique shapes.

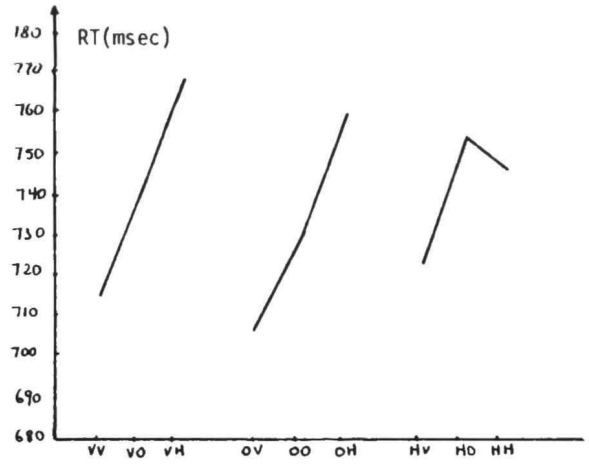


Fig. 5

An ANOVA shows that recognition time depends upon testing orientation only: vertical shapes are faster than oblique shapes and oblique shapes faster than horizontal shapes (F 21, P 10^{-6}, df 140). The effect of learning orientation and the interaction between learning and testing orientations are non-significant. As predicted by the intrinsic encoding model, vertical shapes are easy to recognize, irrespective of the orientation in which they were learned whereas oblique and horizontal shapes take longer.

The paired comparison of all combinations of learning and testing orientations (Student's T-correlated scores test) further supports the intrinsic model and infirms the retinal model. The crucial comparisons are between VO and OV and between VH and HV. The retinal encoding model predicted that they should be equally slow because they involve the same angular shift between learning and testing orientations. This is obviously not the case. OV and HV are as fast as VV and significantly faster than VO (t=2.7,df=111,P=.01) and VH (t=3.6, df=111,P<.01) respectively. These results support the hypothesis that the intrinsic description is stored in memory, irrespective of the learning orientation, and is retreived faster when the shape to be recognized is vertical than when it is oblique or horizontal.

Non-vertical shapes go through the same extra-processing whether the test orientation is the same as the learning orientation or not; OO and HH are slower than OV and HV respectively, the latter significantly so (t=2.04,df=111,P=.04). Thus, even when retinal descriptions match, a shape is recognized on the basis of its intrinsic description.

There is a strong linear dependence of RT's upon the angle between the intrinsic axis of the target and the vertical, for shapes learned vertically and obliquely. (The reason why it does not hold for shapes learned horizontally will not be discussed here.) This suggests that mental rotation might be involved in the encoding of the intrinsic description: the intrinsic description of a non-vertical shape might be obtained by mentally rotating the shape until its intrinsic axis is vertical in the perceptual frame of reference. The average linear slope (1.5°/msec.) is consistent with the rates of mental rotation found in other experiments (Cooper and Shepard, 1973; Shinar and Owen, 1973).

Conclusions

The experiment above shows that the intrinsic axes of shapes play an important role in their mental representation. The intrinsic description is stored in memory and retrieved during recognition irrespective of the orientation in which a shape is seen.

To store the intrinsic description in memory is both economical and effective. The memory representation is unique and allows the shapes to be recognized in novel orientations. I have found in other experiments that the intrinsic description is also used to compare shapes presented simultaneously in different orientations. These experiments also support the hypothesis that mental rotation is involved in the encoding of the intrinsic description: in lateralization studies, the perception of non-vertical shapes shows a right-hemisphere effect.

The intrinsic description appears to be inherent to the mental representation of shapes with intrinsic axes. Experiments in which shapes are presented very briefly and have to be identified among similar shapes in the same orientation presented immediately afterwards, show that oblique and horizontal shapes require longer exposure times to be identified correctly. This orientation effect can be attributed to the extra-processing involved in the computation of the intrinsic axes and/or the intrinsic description of non-vertical shapes. If shapes could be identified on the basis of a retinal description, indentification should be independent of orientation. Palmer (1978) has shown that the descriptions on which recognition is based are high-level, articulated descriptions. The experiments just described suggest that the intrinsic description must be encoded before recognition occurs. This in turn suggests that the only articulated descriptions are relative to the intrinsic frame of reference.

The following model can now be outlined for the processing of two-dimensional shapes with intrinsic axes. I will assume that the mental representation of shapes consists of a series of descriptions in spatial frames of reference, from low-level and local to high-level and articulated. Low-level descriptions are encoded in the retinal frame of reference (the earliest one being the distribution of light on the retina). Intrinsic axes are computed at an early stage of the processing. It has been shown that symmetry can be computed from local descriptions (Julez, 1971) and I have found that perceived elongation is based on principal axes of inertia which can also be computed from low-level descriptions. The position of the intrinsic axis relative to the retinal vertical determines the perceived orientation of the shape. The perceptual frame of reference is then shifted to the intrinsic frame and higher-level descriptions are elaborated within the intrinsic frame. The intrinsic description is stored in memory. It is the description on which recognition is based.

References

Cooper, L.A. and Shepard, R.N. "Chronometric studies of the rotation of mental images," in W.G. Chase (ed.), Visual information processing. New York: Academic Press, 1973.

Dearborn, G. "Recognition under objective reversal," Psycho. rev., 1899, 6, 98-107

Julez, B. Foundations of cyclopean perception. Chicago: Univ. of Chicago Press, 1971

Palmer, S.E. "Structural aspect of visual similarity, Mem. cogn., 6, 91-97, 1978

Rock, I. Orientation and form. New York: Academic Press, 1973

Shinar, D. and Owen, D.H. "Effects of form rotation on the speed of classification: The development of shape constancy," Perc. psychophys., 1973, 14, 149-154

RELATIONS BETWEEN SCHEMATA-BASED COMPUTATIONAL VISION AND ASPECTS OF VISUAL ATTENTION

by
Roger A. Browse
Department of Computer Science
University of British Columbia
Vancouver B.C. Canada

I. INTRODUCTION

This paper explores relations between aspects of visual attention and the operations of schemata-based computational vision systems. These relations are shown to suggest the requirement for methods which operate towards interpretation without model invocation. A specific mechanism is described which permits interpretation based interaction between information from different resolution levels, but does not rely on model invocation. This mechanism is then used in the examination of some related perceptual phenomena, permitting a more computational view of their operations.

II. SCHEMATA-BASED VISION SYSTEMS

An issue of interest to both cognitive psychology and artificial intelligence is the question of how knowledge of a domain of objects can be applied towards visual interpretation. Schemata-based knowledge organizations (Rumelhart and Ortony, 1976; Neisser, 1976) are now being used to address this issue (Freuder, 1976; Havens, 1978; Havens and Mackworth, 1980; Browse, 1980). One distinctive feature of a schemata-based interpretation is the organization of its domain knowledge along "natural" lines. The knowledge is object centered and relies on familiar structuring mechanisms such as component and instance hierarchies.

A domain of knowledge structured in this way is conducive to a recursive cuing mechanism (Havens, 1978): basic image elements act as cues for simple scene objects, which in turn act as cues for more complex objects, etc.

> For example; in the domain of line drawings of human-like body forms (Browse, 1980; 1981), a certain configuration of lines may cue a "hand", which in turn cues "arm", which cues "body".

At each level of this hierarchy, objects are described as being composed of simpler objects.

The occurrence of the objects which are required in the description may not be enough, however, to confirm the existence of the more complex object. There are also relations which must be valid among the components. This distinction will be referred to as the distinction between having found the required elements and having met the required relations.

> For example, all the required elements may exist to make up an "arm": the "hand", the "upper-arm", and the "lower-arm", but a number of required relations must also hold. The elements must be connected in a certain way, and the angles between the elements must be within certain bounds.

While it is difficult to be certain of the presence of an object on the basis of the required elements only, we shall see that there are special situations in which this information is very valuable. These situations rely on a capability of grouping image elements.

During the interpretation process, any element X in the image will have associated with it a set of model possibilities (or labels). This set is simply the set of all objects which are described using X as a required element. In the absence of a means of grouping elements, the interpretation process may deal with the discovery of an element by taking the course of model invocation (or testing). This operation involves selecting one or more of the model possibilities, and testing for their existence by locating the other required elements, and determining the validity of their required relations.

The model invocation approach can provide a dynamic determination of whether the processing procedes top-down or bottom-up (see Havens, 1978). As well it can provide a means of iterative refinement of interpretation and segmentation (see Mackworth, 1978). The operation of model invocation can, however, be costly because it is exhaustive search over the model possibilities.

On some occasions it may be feasible to delete some of the model possibilities without actually invoking them. This is possible whenever uniform constraining relations can be devised over a type of image element.

> For example, if we know that certain lines must be a part of the same object, then the model possibility sets for those lines can be intersected.

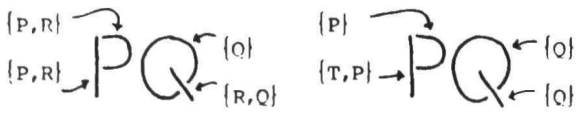
Waltz (1972) has shown that such a uniform constraint may be formulated for the interpretation of line drawings of the blocks-world. The result was that subsequent backtrack search was seldom required. Mackworth (1977) has provided a generalization of the use of such network consistency methods in artificial intelligence problems.

III. FEATURE INTEGRATION AND MODEL INVOCATION

Recently in cognitive psychology there have emerged theoretical ideas about visual attention which relate to the notions of model invocation and operations on sets of model possibilities. The Feature Integration Theory of attention (Treisman and Gelade, 1980) proposes that individual image features are detected rapidly and in parallel, but, in order that an object be identified as consisting of two or more separate features, locations must be processed serially with focal attention. If this is prevented, illusory conjuctions may be formed (Treisman and Schmidt, 1981). Thus, in human vision, the application of focal attention is required by model invocation.

There is an increased expense which accompanies the application of focal attention. Treisman, Sykes, and Gelade (1977) have shown that the amount of time required to detect objects made up of a conjunction of features increases linearly with the display size, but that display size does not have such a great effect on the detection of objects which can be defined without consideration of the relations among features. In computational terms, these objects can be identified only by the examination of model possibility sets to determine the presence of required elements, whereas conjunction objects require the establishment of the relation of common spatial location between features.

To make the relation more clear, consider an example taken from experiment IV of Treisman and Gelade (1980):



(a) target "R" (b) target "T"

Figure 1: two search conditions depicted as features which compose the letters, with their sets of model possibilities attached.

The task was to detect the presence of one of the targets "R" or "I" in a visual field of "P"s and "Q"s. Feature integration theory predicts that the search time will increase linearly with display size for the "R" target, and will increase less for the "T" target. This prediction was found to be correct. The difference between the two conditions is shown in figure 1. For the target "R", the required elements (features) are available in two different ways: either by the presence of an "R" or by adjacent "P" and "Q". In this target condition, the relations among the required elements must be examined. Computationally, this means invoking the model for an "R" each time its required elements are present. The target "T" can be detected by only examining the model possibilities for the primitives because the required features can only be present if the target is present.

There are indications that it is computationally more expensive to invoke models than to use consistency methods (Waltz, 1972; Mackworth, 1977). The proposed relation between model invocation and feature integration adds to the justification for the search for computational methods which can operate towards interpretation at the pre-invocation level.

IV. FILTERING ACROSS RESOLUTION LEVELS

Following the clues provided in the previous sections, Browse (1981) has devised a method which permits the interaction between information obtained at different levels of resolution. This method operates towards interpretation, but before model invocation.

A schemata-based representation for the knowledge of the body-form structure has been developed. This knowledge is specified in terms of image primitives attainable at two different levels of resolution in the image (lines and blobs). Areas of the image for which the correspondence of image primitives across levels is known are areas in which the following uniform constraining relations may be applied:

1. For any blob which must have an integral interpretation, the corresponding lines must all have a common interpretation model. Thus the sets of model possibilities for each of the lines may be reduced to their intersection.
2. The ultimate interpretation must be the same for both levels of resolution (at least instance-hierarchy related), hence the possibility sets may be intersected across levels.

Figure 2a depicts two lines which are known to correspond to a specific blob because of their image hierarchy structure. Also depicted is a model possibility set for each element. By applying rule

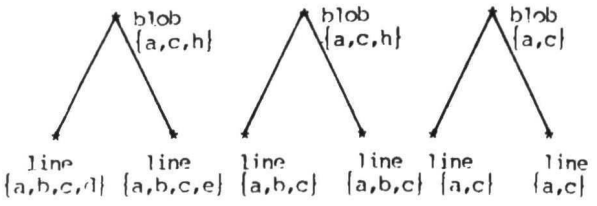


Figure 2: Three stages in the application of consistency across resolution levels.

(1) we eliminate {d,e}, and by applying rule (2) we arrive at the final set of possible models {a,c} as shown in figure 2c. See Browse (1981) for an example of the operation of these methods in the body-form domain.

The correspondences being utilized by these methods are only available in the limited area of the image which has been processed at the highest level of resolution (fovea), and the ultimate usefulness of such operations will be influenced by the appropriateness of the selection of these locations by the program (Browse, 1980). It also remains to establish computational advantages in the order of processing the local and global information (see Navon, 1977; Kinchla and Wolfe, 1979).

V. GROUP PROCESSING AND MODEL INVOCATION

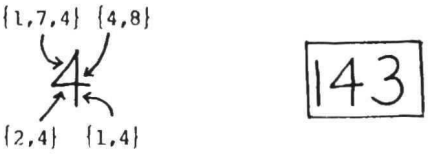
Kahneman and Henik (1977) have formulated a "group-processing" model of the application of attention which is similar to the application of constraining relation (1) of the previous section. Their model proposes a pre-attentive grouping operation which selects large scale objects for subsequent analysis. The experiments which demonstrate the validity of this model employ displays such as that of figure 3.



(a) (b)

Figure 3. Group processing digit detection display

One of the two displays such as shown in figure 3, is presented briefly and the task is to detect a specified target digit. The results show that groups are processed separately, but that processing is almost uniform within groups.



(a) (b)

Figure 4. Features available at two resolutions.

Assume that high resolution feature information is availa[illegible], and that for each such feature, a set of model possibilities is established (as shown in figure 4a). Also assume the availabilty of coarse level

information which gives the identification of larger objects (figure 4b).

Consider the following interpretation of these results: In the first stage, the global objects are detected, as are the high resolution features specifying their model possibility sets. These sets can only be assigned, however, to the established objects, as depicted in figure 5.

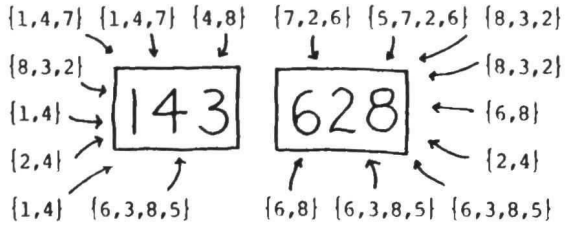


Figure 5. low resolution objects detected and model possibilities assigned to high resolution features, which are roughly located.

At this point, there are obviously too many features associated with the object for it to be a single digit, so a subsequent breakdown of objects takes place. In that this second phase is a higher resolution, it can only take place over a smaller area, so one of the two main objects is selected for more detailed examination (see figure 6).

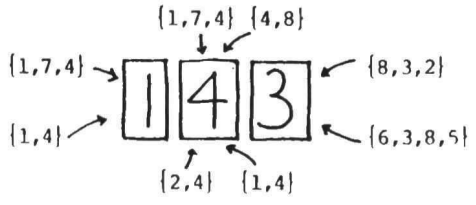


Figure 6. Features assigned to objects detected at a finer level of resolution, for one of the low level objects.

A second examination of the possibility sets reveals that the required elements are available for only one digit in each of the defined positions, and hence their identities can be established in parallel, without serial application of attention to each of the specific locations.

Feature integration theory proposes that object identification may take place in parallel, based on features alone, or serially based on conjunctions of features when necessary. The group processing results indicate intermediate steps at which features are assigned to objects detected at low resolution, and once this assignment is complete, some model possibilities may be discarded by using the constraining relation (1) from the previous section. The location of objects to which these features are attached will become more refined if necessary, to the point of either allowing object identification through confirmation of the presence of the required elements alone, or if necessary by considering the relations among features.

The identity of features may be determined over a wide visual field, but without specific location. Location may become more specific through attachment to low resolution image elements, but only over a more restricted visual field. Finally, the actual location may be determined to permit feature integration. This final locating action operates over a small area of the visual field, and therefore requires serial application if more than one location is to be searched.

VI. SUMMARY

By adopting a schemata-based approach, computational vision domains may be structured so as to use the component hierarchy as a mechanism for cuing the models to be invoked. The complete examination of the relations required by models can be computationally expensive, and for human vision, presupposes the application of foveal attention.

A mechanism has been described which permits the interaction between information from different levels of resolution by eliminating model possibilities and imposing groupings over high resolution features. This mechanism has been shown to be useful in describing the relation between group processing phenomena and Feature Integration Theory.

VII. REFERENCES

Browse, R.A.
1980. Mediation between central and peripheral processes: useful knowledge structures. PROC CSCSI-3, Victoria, B.C. pp.166-171.

Browse, R.A.
1981. Interpretation-based interaction between levels of resolution. submitted to IJCAI-6, Vancouver B.C., Aug 1981.

Freuder, E.C.
1976. A computer system for visual recognition using active knowledge, Ph.D. thesis, AI-TR-345, M.I.T., Cambridge, Mass., 1976.

Havens, W.S.
1978. A procedural model for machine perception. Ph.D. thesis, Thechnical report TR-78-3, Department of Computer Science, University of British Columbia.

Havens, W.S. and Mackworth, A.K.
1980. Schemata-based understanding of hand-drawn sketch maps. PROC CSCSI-3, Victoria, B.C. pp.172-178.

Kahneman, D. and Henik, A.
1977. Effects of visual grouping on immediate recall and selective attention. in S. Dornic (ed.) Attention and Performance VI Hillsdale, N.J. Lawrence Erlbaum. pp.307-332.

Kinchla, R.A. and Wolfe, J.M.
1979. The order of visual processing: "top-down", "bottom-up", or "middle-out". Perception and Psychophysics, 1979,25, pp. 225-231.

Mackworth, A.K.
1977. Consistency in networks of relations. Artificial Intelligence 8, pp.99-118.

Mackworth, A.K.
1978. Vision research strategy: black magic, metaphors, mechanisms, mini-worlds, and maps. in A.R Hanson and E.M. Riseman (eds.),Computer Vision Systems, Academic Press, 1978. pp.53-59.

Navon, D.
1977. Forest before trees: the precedence of global features in visual perception. Cognitive Psychology, 1977,,9. pp. 353-383.

Rumelhart, D.E. and Ortony, A.
1976. The representation of knowledge in memory. Center for Huamn Information Processing report 55, University of California, San Diego. Jan 1976.

Treisman, A., Seykes, M., and Gelade, G.
1977. Selective attention and stimulus integration. in S. Dornic (ed.) Attention and Performance VI Hillsdale, N.J. Lawrence Erlbaum. pp.333-361.

Treisman, A. and Gelade, G.
1980. A feature integration theory of attention. Cognitive Psychology 12 pp.97-136.

Treisman, A. and Schmidt, H.
1981. Illusory conjunctions in the perception of objects. to appear.

Waltz, D.L.
1972. Generating semantic descriptions from drawings of scenes with shadows. Technical report MAC AI-TR-271, M.I.T., Cambridge, Mass.

GROWING SCHEMAS OUT OF INTERVIEWS

Jerry R. Hobbs (SRI International) and Michael H. Agar (Univ. of Maryland)

Ethnography faces complex worlds with no explicit theory. AI, in contrast, carries complex formal theories into encounters with simple worlds. Our work is an effort to find a middle-ground, noting along the way the modifications in both fields required for a synthesis to occur. In this paper we report on our current version of this synthesis by analyzing a fragment of a life history interview with a career heroin addict.

A core problem for ethnographic research is the management of large amounts of qualitative data whose form and content were primarily under informant control. A particular tension in the analysis of this type of material lies in an ethnographer's desire to attend to detail while at the same time offerring more global statements about group life. In research over the last year with an extensive anthropological life history, we have tried different ways to resolve this tension. We would like to report on and demonstrate part of a proposed solution.

The life history analyzed here was conducted over an eighteen month period with an older career heroin addict whom we call "Jack." At the time of the interviews in the early 70's, Jack was about 60 years old, enrolled in a methadone progrm in New York City. The specific interview used for this discussion centers around Jack's story of how he became a burglar. In other papers, we have looked at pieces of this interview to develop our approach. Now we would like to take the interview as a whole to show the interaction between detailed microanalysis of a portion of text and the validation and enrichment of that analysis across the text as a whole. Eventually, we hope to use the approach to treat the entire life history.

Our goal here is to tackle the issue of relating schemas developed in the analysis of a small segment of text to the interview as a whole. We begin with an effort to get a sense of the overall organization of the interview. Our assumption is that the interview, analyzed as a completed act, can be seen as the expression of an informant's plan. We make no assumption that the plan is a representation of what the informant "really" thought, nor do we assume that a plan was consciously worked out in detail before the interview. On the contrary, an earlier paper shows that viewing the completed interview as an expression of a plan forces on us assumptions that highlight the creative emergence of Jack's story.

At the same time, the planning view gives us a sense of the global coherence of the interview, a sense of how different pieces hang together to produce an understanding of the interview as a whole. To get at these "pieces," we first do a high-level segmentation of the interview that makes cuts using major shifts in content as the guideline. Though this process is hardly foolproof, most of the spots for cuts seem intuitively obvious. There is an assumption here that members would mark major segment boundaries in the same way, but we have not tested it out. Further, the ease of segmenting is made easier still by the fact that we are working with data produced by another speaker of American English.

At any rate, once the segments are marked, the problem is to infer the plan of which they are interrelated expressions. Some of the high level goals for an interview are in fact explicitly negotiated in the segments themselves. Where such explicit discussion is not available, we are forced to infer goals and subgoals whose interrelationships provide a coherent account of the interview as a whole. Like most students of phenomena--natural or human--we assume an implicit order that it is our task to bring to light.

The results of the global analysis leave us with a sense of the major segments of the interview together with the goals and subgoals that show them to be coherently linked. The next step is to pick a segment and look for coherence at a lower level--what we call local coherence. This" microanalysis" begins by specifying what it is that each utterance has to do with the ones that immediately precede and follow it. The analysis presupposes that we have a sense of utterance content, a presupposition that is again facilitated here by working with another speaker of American English. The microanalysis in terms of local coherence forces us to specify the logical relations between utterances such that they are seen as coparticipants in connected discourse.

Hobbs has explored the adequacy of a fairly small number of relations to serve in this analysis. The coherence relations, which first organize the utterance by utterance structure in a segment, are then also used to show relations across groups of utterances in that same segment. The coherence relations, in short, are used to show us the ties amont the utterances within a segment.

However, explicating the relationships forces on us the next step in the analysis. If two utterances are related because one "elaborates" on another, we must now make explicit the propositions that justify our claim. If two sequences of utterances in a segment are said to "contrast," we must now show the knowledge in terms of which that contrast can be seen. The local coherence analysis of a segment forces us to develop explicit inferences that make sensible those relations. As will be seen shortly, some of these inferences bunch together through their interlinked predicates and arguments. This "bunching" of inferences, so characteristic of human knowledge, was the reason for the development of the notion of "schema" in AI and psychology. In short, our analysis in terms of local coherence leads us to construct schemas that justify that analysis.

Now, schemas are of particular interest to ethnographers, because they are potentially useful in understanding not just the segment that motivated their construction, but other segments as well. High-level schemas that offer such understanding of a variety of acts have been a traditional goal of cultural anthropology, whether called "patterns," "themes," or "value orientations." However, ethnographers typically construct the high level schemas and demonstrate the resulting understanding in a coarse-grained way. It is this gap that the more detailed local coherence analysis can fill.

At the same time, a local coherence analysis of every segment to which the schema is applied would be too time-consuming. To solve this problem in the sample analysis of one interview presented here, we have developed the following strategy. We pick a particularly interesting schema from the microanalyzed segment and set up some conditions under which it should apply to other segments. In the analysis done here, we are interested in the "arrest" schema, so we decide that any segment that concerns itself with illegal acts will qualify.

The "concern" might be reflected in a single utterance, or it might be the focus of an entire segment. It might be semantically encoded in the utterance, or it might be understood only through inferences connected to that surface semantic content. The segments that satisfy these conditions are then examined for their schematic relevance. As we will show, this process leads to a richer understanding of the details of the schema, a better sense of its relation to other schemas, and validation through its use in understanding other segments of the interview.

However, the examination of the range of application of the schema will not contain the detail of the microanalysis that produced it. We will stop the discussion at the point where we feel that the connection is obvious. At the same time, there is an

assumption that such a detailed analysis is possible for each segment; in principle an analysis of local coherence could be done that would explicitly show the connections. In other words, we will trade off detail for breadth of coverage, without abandoning the obligation to fill in the detail should it be required.

This careful use of different levels of description in different analytical contexts will, we hope, resolve the tension between detailed analysis and breadth of coverage. The strategy is hardly unique to our approach. Learning often works like this--the beginner attends to low-level detail, gradually builds higher level knowledge of what he is doing, and eventually develops a global sense of whatever he is learning and forgets the details unless some problem forces him to return to that level to solve it. We are simply trying to learn to understand an interview in a way that points to strategies for learning to understand even broader ranges of human action.

In the presentation, we will begin by displaying the global plan of the burglary interview. Following that, we will show a microanalysis of a segment to demonstrate the construction of some schemas related to arrest. Then we will look at the schema as it thematically recurs in other segments of the interview and modify and enrich it. Finally, we will conclude with some thoughts on the potential of the method for wider application.

SHAPING EXPLANATIONS:
Effects of questioning on text interpretation

Richard H. Granger, Jr.
Artificial Intelligence Project
Computer Science Department
University of California
Irvine, California 92717

ABSTRACT

Results in cognitive psychology have shown that readers can be steered away from an otherwise plausible interpretation of a story by extra-textual factors such as the source of the text, the stated reading purpose, interruptions and repetition of questions about the text. For instance, successive repetitions of the same question about a given text will often elicit a series of alternative interpretations of the text. This effect cannot be accounted for by established principles of text processing behavior, such as people's preference for cohesive and parsimonious representations of text. This paper presents a computer program called MACARTHUR, which models this behavior by varying the depth and direction of its inference pursuit in response to re-questioning, resulting in a series of markedly different interpretations of the same text. In light of the results, some new experiments are suggested in hopes of arriving at a new principle, beyond cohesion and parsimony, to account for the observed text processing behavior.

1.0 INTRODUCTION

Consider the following story:

[1] The Pakistani Ambassador to the United States made an unscheduled stop in Albania yesterday on his way home to what an aide of the Ambassador described as "a working vacation".

Why did the ambassador go to Albania? People in informal experiments most often answer that he may have simply gone there as part of his vacation. However, when the same question is repeated, they generate alternative explanations, such as the following:

1. There could have been some secret political meeting there.

2. There might have been plane trouble; say, an emergency landing to fix a fuel leak.

3. Maybe he just wanted to avoid reporters on his vacation.

The text presents an explanation on the surface (that the ambassador was on vacation), which is adequate to serve as an interpretation of the events in the story. However, readers can be steered away from this explanation by external factors such as repetition of the same question. In a related series of informal experiments, people were told different "sources" of the text; in particular, they were either told that it was excerpted from the New York Times, an Agatha Christie novel, Cosmopolitan magazine, a grammar-school history textbook or a Jimmy Stewart movie. Their interpretations of the text varied significantly depending on the stated text source.

These observations about people's reading behavior agree with experiments in cognitive psychology in which varying the stated reading purpose (e.g., Black [1980], Frederiksen [1975]), and interposing questions about the text (e.g., Rothkopf and Bisbicos [1967], Anderson and Biddle [1975]) resulted in differences in inferences made by the readers, as evidenced by tests for false recognition of statements corresponding to inferences from the text.

This paper presents a program called MACARTHUR which is able to redirect its own inference processes when a question about a text is re-asked repeatedly. MACARTHUR demonstrates its successive interpretations by generating English answers to questions about the text. For example, after reading a version of the above story [1], MACARTHUR responds in English to the following sequence of questions:

Q) Why did the ambassador go to Albania?
A) HE WENT ON A VACATION IN ALBANIA AND PAKISTAN.

Q) Are you sure? Why did he go to Albania?
A) MAYBE HE WANTED TO MEET WITH THE GOVERNMENT OF ALBANIA, BUT HE WANTED TO KEEP IT A SECRET.

Most existing text understanding systems (e.g., Cullingford [1978], Wilensky [1978], DeJong [1979], Charniak [1978]) do not account for people's ability to make different inferences depending on external factors such as re-probing. MACARTHUR's ability to re-direct its own inferences arises from a new classification scheme for explanations based on an attribute termed the "shape" of an explanation. The program is intended to provide a test-bed for comparing implementations of our theories about people's reading behavior with actual experimental evidence. Towards this end, the concluding section of this paper proposes some possible new experiments, and some possible extensions to MACARTHUR.

2.0 BACKGROUND: COHESION AND PARSIMONY

2.1 The cohesion principle

Results in cognitive psychology have shown that people almost universally construct interpretations of text which serve to coherently connect the separate statements in a text, even when such connections are not at all obvious. For instance, Haberlandt and Bingham [1978] have found evidence for causal connective inferences being made among the sentences in examples like the following:

[2] Brian punched George. George called the doctor. The doctor arrived.

[3] Brian punched George. George liked the doctor. The doctor arrived.

Subjects took longer to read [3] than [2], presumably spending the extra time trying to infer causal or intentional connective inferences among the statements in the text.

Similarly, Bower, Black and Turner [1979] found that reading times were longer when readers had to perform more than one inferential "step" to establish a causal connection between two statements in a narrative. These results and others providing evidence for spatial, instrumental, referential, causal and intentional connective inference have demonstrated that a crucial feature of human text understanding is the ability to construct a connected and coherent representation of a text. Taken together, these results form what we may term the "cohesion principle" of text processing behavior.

Researchers in AI have constructed a number of process models of text understanding which are consistent with the cohesion principle. AI programs that have addressed the problem of connectedness in texts include the MARGIE program (Schank [1975]) in terms of causal connections, the SAM and Ms.Malaprop programs in terms of script- and frame-based connections (Schank et al [1975], Cullingford [1978], Charniak [1979]), and the PAM and BELIEVER programs in terms of intentional connections (Schank and Abelson [1977], Wilensky [1978], Sridharan and Schmidt [1978]).

2.2 The parsimony principle

The cohesion principle alone is not sufficient to account for people's interpretations of text. For instance, consider the following deceptively simple example (from Granger [1980]):

[4] Mary picked up a magazine. She swatted a fly.

When asked why Mary picked up the magazine, people in informal experiments overwhelmingly answer that she picked it up with the intention of swatting the fly. However, this answer corresponds to only one of (at least) three possible interpretations of the text, none of which can be ruled out on grounds of logic or the cohesion principle:

(4a) Mary picked up a magazine to read it. She then was annoyed by a fly, and she swatted it with the magazine she was holding.

(4b) Mary picked up a magazine to read it. She then was annoyed by a fly, and she swatted it with a flyswatter that was handy.

(4c) Mary picked up a magazine to swat a fly with it.

This same phenomenon occurs in any "garden path" text; i.e., a text that suggests an initially plausible inference which turns out to be "supplanted" (Granger [1980]) in the final representation. To account for these observations, Granger proposed the Parsimony Principle, which states that the preferred interpretation of a text is the one in which the fewest number of inferred intentions of a story character account for the maximum number of his actions. This principle has been incorporated into a computer program called ARTHUR (A Reader THat Understands Reflectively), which can supplant its own initial inferences in light of subsequent information in a text, thereby enabling it to read garden path stories.

3.0 THE SHAPE OF EXPLANATIONS

The cohesion and parsimony principles together still fail to account deterministically for certain text understanding behavior. In particular, people's ability to generate alternative interpretations of a text in response to re-questioning cannot be explained by these principles, since, for example, all four of the interpretations given earlier in this paper for story [1] are coherent and parsimonious.

In order to account for this behavior, we have developed a classification scheme for alternative explanations based on an attribute of explanations we term their "shape". This scheme has proven useful in the explanation-selection algorithm used by MACARTHUR in generating alternative interpretations of a text. Following is a list of the four shapes MACARTHUR currently knows about. This is not intended to be a complete list, it simply reflects the present state of our analysis:

1. Pursue-desired-state: This refers to simple goal pursuit, i.e. a story in which a character has a goal and performs plans in service of that goal.

2. Avoid-undesired-state: A character may not have a specific goal or desired state, but rather is acting out plans that are in service of the avoidance of a particular undesired state, such as sleepiness (for which a remedy is to ingest coffee or other stimulants), hunger (remedies include doing something distracting like reading, or taking diet pills, or even going to sleep), etc.

3. Accident-reaction: A character may be involved in some events that unintentionally hinder his goals. The character's subsequent actions may include attempts to investigate the cause of the accident; overcoming the accident by re-planning and reacting; abandoning or postponing the goal; or simply trying again.

4. Cover-stories: A character may have a goal that he wishes to achieve secretly. If he cannot simply avoid being observed, then he may construct a "cover story"; i.e., an alternative connected explanation for his actions which can serve as an "alibi" to any observers. Complete understanding of such stories involves the ability to maintain separate belief spaces for different characters, and to recognize deception via conflicting beliefs held by different characters.

Following is an illustration of how these explanation shapes can give rise to a series of alternative interpretations of stories. Recall story [1]:

[1] The Pakistani Ambassador to the United States made an unscheduled stop in Albania yesterday on his way home to what an aide of the Ambassador described as "a working vacation".

The four alternative explanations previously given for this story can now be categorized by explanation shape:

1. He may have gone there as part of his vacation. (PURSUE-DESIRED-STATE)

2. There could have been some secret political meeting there. (COVER-STORY)

3. There might have been plane trouble; say an emergency landing. (ACCIDENT-REACTION)

4. Maybe he just wanted to avoid reporters on his vacation. (AVOID-UNDESIRABLE-STATE)

Consider story [5], another story that MACARTHUR can process (see Granger [1981] for examples of detailed output from MACARTHUR):

[5] Dr. Fitzsimmons yawned loudly. He left Carney and Samuelson and went into the next room. He opened the refrigerator.

Following are four differently-shaped explanations for this story.

1. Maybe he wanted to make some warm milk to help him get to sleep. (PURSUE-DESIRED-STATE)

2. Maybe he wanted to make some coffee to help him stay awake. (AVOID-UNDESIRABLE-STATE)

3. Maybe he heard something fall down in there and he went to investigate. (ACCIDENT-REACTION)

4. Maybe he actually had some secret reason for going in there, so he yawned to pretend he was tired. (COVER-STORY)

4.0 CONCLUSIONS: PROPOSED EXPERIMENTS

Black's [1980] experiments on the effects of reading purpose on memory for text assumed that the task of rating the comprehensibility of a text was "a 'shallow' task", preparing for a memory test was "a 'deeper' task", and preparing for an essay test in which the subjects would have to make use of the main point of the text was "a 'deepest' task" [p. 20]. Black's initial prediction was basically that the "deeper" the reading purpose, the greater the number of inferences the subject would produce, as evidenced by the number of false recognitions exhibited on tested inference items.

The actual results of the experiment indicated that the memory task caused the most false recognitions of inference items, while the essay task came second and the comprehensibility task came lowest, as expected. A post-hoc analysis of the recognition test items revealed that the essay task caused significantly more false recognitions than the other two groups on inference items which were "related to the main point" of the story, even though the number of false recognitions overall (i.e., including items both related and unrelated to the main point) was lower for the essay task than for the memory task.

In other words, the experiment was looking for a monotonically increasing effect of more inferences corresponding to "deeper" processing. However, what it found was a difference in not only the "depth", but also in the "direction" of inferences generated. In particular, Black acknowledges the existence of "main-point oriented" processing in the essay task which did not appear in the other two tasks.

Consider a similar set of experiments based on more difficult stories, i.e., stories that are less strongly connected to a single main point than the essays used in Black's study. For example, non-straightforward texts like [1] and [5] in this paper could be used. According to the cohesion principle, readers tend to work at finding connections among sentences in a text, even when such connections are not obvious. Hence, we predict that subjects would dutifully generate connective inferences to explain the sentences in these non-straightforward texts. However, since there are a number of different alternative interpretations for these texts, different explanations might be produced by different subjects, perhaps as a function of different types of external factors such as reading purpose, text source, interposed questions and re-probing. For example, in a reading-purpose experiment the "shallower" readers might generate a "naive" interpretation of a difficult text; while deeper readers might generate not just more inferences but different inferences, corresponding to their significantly different interpretation of the text. We propose such a set of experiments, designed around non-straightforward texts, and making use of other types of extra-textual factors than just reading purpose; in particular, the effects of interposed questions and re-probing.

REFERENCES

[1] Anderson, R.C. and Biddle, W.B. (1975) On asking people questions about what they are reading. In G. H. Bower (Ed.) The psychology of learning and motivation. Vol 9. Academic Press, New York.

[2] Black, J. (1980). The Effects of Reading Purpose on Memory for Text, Cognitive Science TR #7, Yale University, New Haven, Conn.

[3] Black, J. and Bern,H. (1980). Causal Coherence and Memory for Events in Narratives, Cognitive Science TR #3, Yale University, New Haven, Conn.

[4] Bower, G.H., Black, J.B. and Turner,T. (1979). Scripts in memory for text., Cognitive Psychology, 11, 177-220.

[5] Charniak, E. (1978). On the use of framed knowledge in language comprehension. Computer Science TR #137, Yale University, New Haven, Conn.

[6] Cullingford, R. (1978). Script Application: Computer Understanding of Newspaper Stories. Ph.D. Thesis, Yale University, New Haven, Conn.

[7] Frederiksen, C.H. (1975). Effects of context-induced processing operations on semantic information acquired from discourse. Cognitive Psychology, 7, p. 371-458.

[8] Granger, R. (1980). When Expectation Fails: Towards a self-correcting inference system. Proceedings of the First National Conference on Artificial Intelligence, Stanford, Cal.

[9] Granger, R. (1981). Directing and Re-directing Inference Pursuit: Extra-Textual Effects on Text Interpretation. Technical Report #171, Computer Science Dept., University of California, Irvine, Cal. Submitted to the 7th IJCAI, Vancouver.

[10] Haberlandt, K. and Bingham, G. (1978). Verbs contribute to the coherence of brief narratives: Reading related and unrelated sentence triples. Journal of Verbal Learning and Behavior, 17, p. 419-425.

[11] Rothkopf, E.Z. and Bisbicos, E. (1967). Selective facilitative effects of interspersed questions on learning from written material. Journal of Educational Psychology, 58, p. 56-61.

[12] Schank, R. C. (1975). Conceptual Information Processing, North-Holland, Amsterdam.

[13] Schank, R. C. and Abelson, R. P. (1977). Scripts, Plans, Goals and Understanding. Erlbaum Press, Hillsdale, N.J.

[14] Sridharan,N.S. and Schmidt,C.F. (1978). Knowledge-directed inference in BELIEVER. In Waterman & Hayes-Roth (Eds.), Pattern-Directed Inference Systems, Academic Press, N.Y.

[15] Wilensky, R. (1978). Understanding Goal-based Stories. Ph.D. Thesis, Yale University, New Haven, Conn.

The Need for Context in Event Identity

John M. Morris
4 Proctor Avenue
Clinton, New York

The problem of event identity is the problem of determining whether two sentences describe the same event.

Recent discussions have suggested two possible criteria for determining event identity: (1) identical times, objects, and constituent properties; and (2) identical causes and effects. Both approaches can be shown to be subject to fatal exceptions. In everyday life, however, we have little trouble in determining what specific criteria would be relevant to determining event identity. I want to show that such criteria are heavily context-dependent, suggesting that event identity cannot be determined without consideration of the context in which the event occurs.

Here are some examples of the ways in which one event may be said to be identical to another:

A. Logical Entailment

A1. The plane flew over the UN Building, and the UN Building is in New York.

A2. The plane flew over New York.

B. Definitional

B1. Carol forced open the window, reached through the window, and removed the depression-glass vase, with the intention of appropriating it to her own use.

B2. Carol committed the crime of unlawful entry.

C. Conventional

C1. He shouted "Au secours."

C2. He cried for help.

D. Causal

D1. She struck her leg sharply against the fence rail.

D2. She broke her leg.

E. Accidental

E1. A hero of the Resistance was elected French president.

E2. A man with a big nose was elected French president.

F. Psychoanalytical

F1. A child dreamed of a dragon.

F2. A child dreamed of her father.

G. Theoretical

G1. The water boiled.

G2. The molecules moved rapidly.

H. Epistemic

H1. Spring is coming.

H2. Days are getting longer.

There is nothing particularly sacred about this set of categories, and you are welcome to expand, contract, modify, or otherwise mutilate the list at will. The important fact about the list is simply that it illustrates the variety of ways in which two sentences can both assert that the same event occurred.

The logical entailment in Category A is, of course, rather loose, but the point is simply that (given appropriate background knowledge) A1 provides completely conclusive evidence for A2; one cannot deny A2 without (implicitly) denying A1. We know this because we know the meaning of such words as "flies" and "over"; no searches for empirical evidence could possibly substitute for this elementary understanding.

In category B, Definitional, the meaning is scarcely part of our everyday knowledge of the language, but requires a search through a legal dictionary. To say that Carol committed the crime is to say that she performed a certain set of acts, and it is one of the purposes of the law to define these acts as precisely as possible.

Notice in particular that Carol's reaching through the window and appropriating the vase was not a different event from her committing the crime. We should scarcely want to say that she committed the crime and <u>also</u> reached through the window. Instead, what we say is that she committed the crime <u>in</u> reaching (or <u>by</u> reaching) through the window, where the reaching occurred within a particular context of property ownership and malicious intent. The criteria that identify the event as a crime are part of the context in which it occurs.

The next category, Category C, is full of interest for the philosophy of language, and I will only hint at it here. Suppose that at a private beach, a drowning person shouts "Au secours!" The lifeguard, not understanding, fails to respond. The bereaved family sues the beach, on the ground that the guard failed to perform his obvious duty. Let us suppose that the case resolves itself to the issue of whether the man did or did not call for help. Under what conditions would we decide for the lifeguard or for the family? The problem is (I think) extremely difficult and interesting, but it is not the sort of problem that can be solved by looking in the dictionary or by gathering evidence; it depends on how the language is to be used.

Category D is called "Causal," and I have included this example simply to show that causes are not always prior in time to their effects. She has broken her leg in the very act of hitting it against the fence rail; she did not hit the rail and <u>then</u> break her leg. Nevertheless, if asked, she would be likely to say, "I broke my leg <u>by</u> hitting it against the fence rail." (She certainly would not

say, "I hit my leg by breaking it.") Yet there is nothing to distinguish the hitting-event from the breaking-event; they are the same event as seen in two different ways.

Another category of interest, Category E, Accidental, shows the importance of context in determining how an event will be described. A reasonably patriotic French historian would surely choose E1 over E2 as a way of describing a certain French election. For the historian, E1 might be expected to function in an interesting way in the formulation of various generalizations about voting behavior. This historian might wish to formulate some general rule on the way in which former heroes tend to win elections. The description E2, however, seems totally devoid of interest for the historian, since it is unlikely that any generalization of the slightest historical interest can be derived from it.

On the other hand, we might imagine an overly self-conscious person with a large nose, who is preparing a history of the famous large-nosed people of the world, and who imagines that this election has elevated the status of big-nosed people everywhere. For such a person, description E2 is of the greatest importance.

But surely E1 describes the same event as E2. DeGaulle was a hero of the Resistance, and he was also a man with a big nose; he was not two persons. It was precisely the same election that elevated the hero and the big-nosed man; E1 does not describe a different election from that described by E2. In fact, the most useful way of characterizing the relationship between E1 and E2 is simply to say that they both describe the same event.

The following objection lies at the center of our problem, and it illustrates the confusion with which the problem has been enveloped. It will be objected that the consequences of E1 are obviously different from the consequences of E2.

What sort of consequences might we have in mind? The consequences of E1 would be those of electing a hero -- a surge of patriotism, resentment on the part of anti-war factions, and increase in the influence of the military forces. The consequences of E2 would obviously be quite different -- derision from unfriendly foreign powers, jubilation among political cartoonists, renewed literary interest in Cyrano de Bergerac. Yet the two events are one, because DeGaulle was just one person.

The identity in Category F suggests simply that a psychoanalytically-inclined interpreter might wish to say of the child that she was "really" dreaming of her father. The meaning of "really" will, of course, depend on how we interpret the claims of psychoanalysts and others who believe that our dreams are symbolic of our waking life.

In the pair of examples G1 and G2, the boiling of the water was said to be identical with the movements of its molecules. The type of identity suggests is what I will call "physical" identity, since we would scarcely want to say that the boiling of a pan of water was a different event from the rapid movement of its molecules. At the same time, the context in which a cook is watching the pan of water, to determine when to add the noodles, is quite different from the context in which an early physicist might have watched it, to determine the local turbulence that reveals the molecular agitation. Thus the event in the physicist's context is conceptually quite different from the event in the cook's context.

The goal of the seventeenth-century physicist was to establish that the boiling event (as observed by the cook) was identical to the molecular-agitation event (as predicted by the theory). Since the latter event could not be observed directly, it was necessary to develop techniques for inferring the molecular motions from observable events, such as local turbulence in the water. The reason for doing so was that a great many interesting consequences could be derived from the general theory of molecular motions, which could not be derived from the cook's account. We may say, then, that boiling is physically identical with molecular agitation, but that it is conceptually different from it. This distinction is one which points a way to a solution.

Category H, Epistemic, contains an interesting type of identity which could not conveniently be fitted into any of the others. We know that spring is coming, in the same way that we know that the days are getting longer -- by looking at the calendar or by timing the sunsets. Evidence for H1 is the same as evidence for H2.

But there is no simple way in which H1 can be said to imply H2, or vice-versa. There seems to be no logical relationship between them, as there is between A1 and A2. At the same time, we could hardly say that the approach of spring is different from the lengthening of days; H1 and H2 represent two ways of looking at the same set of happenings.

Thus we have eight different ways of determining the identity of eight different sorts of events. How much do they have in common? I want to give a brief answer to a complex question, with the assurance that a great deal remains to be said.

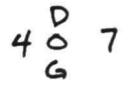
An event -- like a thing or an item -- is defined within a given context of discourse. We count the items on my bookshelf in one way if we are book dealers, in other ways if we are wastepaper collectors or interior decorators. We count the things

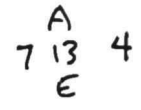
in a field in one way if we are farmers, another way if we are botanists. Similarly, we count events in one way if we are interested in aircraft flights, another way if we are concerned with noise polution or destruction of the ozone layer.

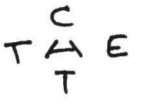
Puzzles occur when we cross the boundaries that separate contexts -- when Eleanor suddenly discovers that the intruder at whom she's been aiming the pistol is really her husband. The event -- his coming into the room unannounced -- is an intrusion in one context, a homecoming in another. Was his entry into the room one event or two? In some sense, it seems possible to say that it was both.

Consider:

J1.


J2.


J3.


J4.


In J1, the center symbol is unambiguously an "A". In J2, the "O" is a figure in which ambiguity depends on whether it is interpreted as a numeral or as a letter. When we read left-right, it is a zero; when we read top-down, it is an oh. In J3, the center figure is either one letter or two numerals, and again the ambiguity is resolved as we read top-down or left-right. In J4, the center symbol is hopelessly ambiguous in isolation. It can be either an "A" or an "H". Ambiguity is resolved as we read top-down or left-right.

We can say that "A" (i.e. the token) in J1 is the same letter (i.e. the same token of the letter "A") whether read top-down or left-right; in either word, the center "A" is numerically identical with the center "A" of the other.

But in J2, it seems strange to say, analogously, that the central numeral "0" is identical with the letter "O", since a numeral is not the same as a letter. The puzzle is not a paradox, however, since we can resolve it by saying something like: "The letter 'O' has the same shape as the numeral '0'." (I do not mean to claim that this is really a very satisfactory resolution of the puzzle; I want only to emphasize that there must be some way of resolving it.)

In J3, the situation seems a little stranger, since one letter appears to be identical to two numerals. Finally, in J4, a token of a letter which we identify as an "A" is identical to a token of a letter when we call an "H", yet we do not want to say that an "A" is identical with an "H".

By insisting that various kinds of identity must be separated out, I have tried to suggest that the problem of event identity is not as difficult as it appeared at first. This approach means that the very nature of an event depends on the context in which it occurs (just as the shape 13 may be either a "B" or a "13", depending on its context).

One way of emphasizing the context-dependency of events is to consider non-events -- that is, those occasions on which an event fails to occur. The watchman fails to make his rounds. The bridegroom does not make it to the altar. There is (I think) a scene in Chekhov in which the young man has been expected to propose to the daughter. His proposal is extremely important to the family, because it promises them a way out of their poverty and debts, thanks to his money and status. He appears at their home, and he plays cards with the family. This is the only event that happens on the stage -- the actors play a perfectly ordinary game of cards. But the family -- and the audience -- know that he has failed to propose. His failure, his cowardice at the crucial moment, means bankruptcy and disgrace for the family, and it is this knowledge that gives the scene its emotional significance.

But a non-event like this is literally nothing. No relevant physical events occur; yet (in another sense) a disaster has occurred for the family. The non-event is like the null set, in that it cannot be distinguished from any other non-event, if they are considered out of context. A small blank area on a sheet of paper may be physically identical with that same blank area when the paper is filled with writing; but it will serve different functions, depending on where the blank space occurs in that writing. The context makes all the difference.

In the series of stories and analogies that I have presented here, I have suggested that ordinary people have rough-and-ready ways of answering questions about event identity, and that we can make some operational sense out of their rough-and-ready methods. The primary method that I have recommended is to draw a sharp distinction between physical identity and conceptual identity, rather than treating event identity as though it were a single type of identity. I have also suggested that conceptual identity will depend on the context in which the event occurs.

REFERENCES

Donald Davidson, "Actions, Reasons, and Causes," Journal of Philosophy, Nov. 7, 1963, pp. 685-700.

Jawgwon Kim, "Causation, Nomic Subsumption, and the Concept of Event," Journal of Philosophy, Apr. 26, 1973, pp. 217-236.

John M. Morris, "Non-Events," Philosophical Studies, 1978.

Point of View in Problem Solving

Edwin L. Hutchins
James A. Levin
Laboratory of Comparative Human Cognition
University of California, San Diego

Problem solvers adopt "points of view" when solving problems, expressed through the deixis of their verbalizations, that are strongly related both to the commission of illegal moves and to the occurrence of blocked conditions. This paper describes an analysis of point of view in the Missionaries & Cannibals task, and presents a model for problem solving that incorporates point of view as a resource allocation mechanism useful for dealing with the the finite capacity of human problem solving processing. This analysis relates the subjects' actions in this task to their talk about these actions.

When solving a problem, where do people put themselves? After solving a simple puzzle (the Missionaries & Cannibals task), all of our subjects reported having taken a "bird's eye view", looking down on the puzzle elements from above. Yet their verbal reports of puzzle actions were in terms of motion relative to their own positions, thus placing the problem solvers within the local space of the problem elements rather than removed from it. The English language permits a speaker to describe motion in space relative to his/her own position (or relative to other spatial landmarks). This problem solving "point of view" shifts over the course of solving the problem. More importantly, a person's point of view is related to progress in solving the problem.

The analysis of what people say while solving problems has played an important role in problem solving research. There has been controversy over the status of what people say about what they are doing. At one end of a spectrum, the manifest content of problem solving "protocols" is taken as an accurate reflection of a subset of problem solving processes (Newell & Simon, 1972). At the other end of the spectrum, this kind of talk about action has been rejected as valid data (Nisbitt & Wilson, 1977).

In this paper we take a position that neither assumes a simple link nor dismisses talk but instead closely looks at what the relation is. A detailed examination of the talk about actions in solving a simple puzzle reveals a systematic relation of which the subjects themselves (and previous researchers studying this puzzle) are not consciously aware. This relation can serve as an important building block of a model of problem solving that encompasses both talk and task actions.

In our experiments, the subjects sat facing the experimenter with the puzzle pieces on a table between them. The puzzle pieces consisted of a piece of paper with a river drawn on it, three tokens labeled with Ms to stand for missionaries, three tokens labeled with Cs to stand for the cannibals, and a paper boat that would hold a maximum of two tokens. The object of the Missionaries & Cannibals puzzle is to get all the people across the river using the boat, without ever having more cannibals on a side than missionaries. In all cases, the verbal interaction between the experimenter and the subject was tape recorded.

Point of View

The subjects represent the spatial aspect of the problem in their accounts primarily through the use of deictic verbs (come, go, take, send, bring, etc) and place adverbs (here, there, across, etc). The use of these lexical items positions the speaker relative to a spatial field.

For example, one subject began the task with the following statement:

"I want one cannibal and one missionary, and they go to the other side, and the guy drops off the cannibal and the missionary comes back again."

The condition which has to be met in order for "go" to be appropriate is that the speaker is not at the goal of the action at the time of the utterance (Fillmore, 1974; Clark & Garnica, 1974). For the verb "come" on the other hand, the condition which must be met is that the speaker is at the goal of the action. In this case we therefore assume that the subject has an implicit point of view on the problem which places him on the start shore throughout the two moves described.

Another subject started the task with the next statement:

"First thing I want to do is get a cannibal over to the other side. Let's take him over there with a missionary. Missionary takes the boat back."

In this case the problem solver has expressed a shifting point of view. At the outset, the subjective point of view of the problem solver is at the start shore. This is shown both by the fact that the verb "take" indicates that the subject is not at the goal of the action at the time of the utterance and by the reference to the goal side of the river as 'the other side.' In the course of the move the point of view changes to the goal shore as the problem solver travels with the creatures in taking them to the other side. The point of view of the problem solver remains at the goal shore through the execution of the next move. This is indicated by the deixis of the phrase 'Missionary takes' which again places the subject on the shore of the origin of the action rather than at the goal of the action.

In previous approaches, (Thomas, 1974; Greeno, 1974; Jeffries, Polson, Razran, & Atwood, 1977) the subject was notified immediately upon the production of an illegal state. In the procedure employed here, illegal states were noted by the experimenter, but the subject was not told that an illegal state had been produced until a following move was attempted. This provided the subject with an opportunity to self detect illegal states. If the subject failed to notice an illegal state, it was pointed out by the experimenter when the next move was attempted. This procedure permits us to distinguish illegal states that are self-detected by the subject from those that go unnoticed by the subject.

Any move can be classified in terms of its actual legality and its judged legality. This classification is shown in the two by two table below.

		JUDGED: legal	JUDGED: illegal
ACTUAL	legal	LEGAL MOVE	BLOCKED CONDITION
	illegal	ILLEGAL MOVE	CORRECT REJECTION

Errors of commission

Of these types of moves, the analysis of moves that produce actual illegal states is the most straightforward, so we will begin with it. Since an illegal state is produced when the cannibals outnumber the missionaries on either side of the river, and since they cannot simultaneously outnumber them on both sides, illegal states have a sidedness relative to the river. Where it is possible to determine the subject's point of view at the time of the illegal move, the illegal state can be labeled a near side illegal state or a far side illegal state. Near side illegal states are those in which the rule violation occurs on the same side of the river as the subject's current point of view. Far side illegal states are those in which the rule violation occurs on the side of the river away from the subject's current point of view.

The results of this analysis is shown in Table 2. Of the 15 illegal moves for which it was possible to assign an unambiguous point of view, 10 occured on the river bank away from the point of view of the subject, while only 5 occured on the river bank of the subject's point of view. Further, four of five near side illegal moves were detected by the subject before making another move, while eight of ten far side illegal moves went undetected by the subject.

	Violation side: Near	Far
Detected by subjects	4	2
Undetected by subjects	1	8
Total	5	10

Errors of Omission

The analysis of errors in problem solving has largely focused on errors of commission, the illegal moves that subjects make. However, a "problem" is not just a situation where a person makes illegal moves. It may also be a situation where a person is unable to progress toward some goal, even after repreated attempts. This situation can be caused by "errors of omission", where the person fails to make a progressive move, as well as by the commission of illegal moves.

Legal moves do not have sidedness in the same way that illegal moves do. As noted above, an illegal state results from a rule violation that is located on one side of the river or the other. When there is no rule violation, there is no sidedness. However, these moves are still amenable to point of view analysis.

Novice problem solvers are sometimes blocked several times at the same state before sucessfully getting through it. These several passes through the same state may show changes in point of view. A particular point of view on the problem may lead the subject to discard a legal move, while a different point of view on the same state may make the legality of the next move obvious.

Early research on problem solving, especially that by the Gestalt psychologists, focussed on obstacles to achieving goals. For example, some of the early work by Kohler (1925) looked at how various organisms dealt with a physical block, a wire mesh fence between the organism and some food. How can we characterize the condition of being blocked? Kohler's animals were blocked when they made repeated attempts to get to the goal, none of which made progress toward the goal. These non-progressive moves included backing away from the fence and running into the fence.

By analogy, we can extend this criterion for being blocked into a more abstract task such as the Missionaries and Cannibals puzzle. A subject is blocked in a state when s/he makes at least two non-progressive moves out of that state with no intervening progressive move from that state.

With this definition of a "blocked condition, we identified fourteen instances in our data across the three experiments. In four instances, the subject expressed a definite "point of view" for both the first move taken when blocked and then the first progressive move that broke through the block. In all four cases, the point of view expressed when blocked was different than when not blocked.

Toward An Activation Model of Problem Solving

We have been developing a dynamic interactive model of problem solving, based on an activation framework for cognitive processing (Levin, 1976, 1981). Within this framework, the current state of the problem solving is modelled by the current set of activations of concepts in the problem solver's long term memory. Each activation influences other activations, increasing or decreasing the salience of its neighbors. Processing resource in this framework is directly captured by the salience metric, as highly salient activations have a large influence on the global result of processing, and activations that lose salience disappear from the scene. More salient activations of concepts are more likely to have effect than less salient contradictory activations.

In this model, possible moves in a problem are activated by their pre- and post-condition states. The current state of the problem will be strongly activated by perception, and actions for which the current state is a pre-condition will thus be salient. Post-conditions that are

similar to the goal state are also more salient. Post-conditions that are "illegal" are inhibited by the constraint concepts. The interaction between current state, goals, and constraints creates a dynamic set of activated moves with differing relative saliences.

Point of view, in this framework, is a salience allocation mechanism, contributing salience to the activations associated with the location of the problem solver's diectic position. The likelihood of an illegal move resulting from a violation of a constraint is inversely related to the salience of the constraint activation on the side where the error occurs.

Illegal Moves. Illegal moves are, in this framework, more likely to occur on the "far" side (away from the point of view position of the solver), since those constraints are less salient than "near side" constraints. In addition, detection of an illegal move once made is more likely when the constraint concepts are more salient.

Blocked conditions. A blocked condition results when the progressive move in a situation is less salient than alternative moves. In the simplest case, the progressive move never acquires enough salience to be activated at all. In this case, the problem solver is totally "unaware" of the progressive move. In a more complicated case, the progressive move is considered, but not taken because it is less salient than alternative moves. A change in point of view may shift the relative saliences of the various simultaneously active alternative moves, and thus can lead to the solver selecting the previously rejected move, surmounting the roadblock to progress.

The appearance of point of view in problem solving protocols and its apparent relation to problem solving processing casts new light on the relation of verbal protocols to the processing they describe. Much of the processing that goes into our problem solving is transparent to the solver. That is, we do it and are not aware that we have done it. In the case of point of view, we not only do it, we speak about it as well, and still we are not aware that we have done so. In fact one could (and many have) read the protocols many times and never notice the use of deixis. These transparent processes are important in our problems solving, but they are mercifully invisible to us. Were they continually in consciousness, we would surely become confused. In analysis, we have the luxury of being able to examine both what is being done and how it is being done. In the phenomenon of point of view in problem solving, we see an aspect of the problem solving processing finding expression in the verbal protocol, without the problem solver being aware of it.

References

Clark, E. V., & Garnica, O. K. Is he coming or going? On the acquisition of deictic verbs. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 559-572.

Ericsson, K. A., & Simon, H. A. Verbal reports as data. Psychological Review, 1980, 87, 215-251.

Fillmore C. Pragmatics and the description of discourse. In C. Fillmore, G. Lakoff, & R. Lakoff (Eds.), Berkeley studies in syntax and semantics. Vol. I. Berkeley, CA: Department of Linguistics, University of California, 1974.

Greeno, J. G. Hobbits and orcs: Acquisition of a sequential concept. Cognitive Psychology, 1974, 6, 270-292.

Jeffries, R., Polson, P. G., Razran, L., & Atwood, M. A process model for missionaries-cannibals and other river-crossing problems. Cognitive Psychology, 1977, 9, 412-440.

Kohler, W. The mentality of apes. London: Routledge and Kegan Paul, 1925.

Levin, J. A. Continuous processing with multilevel memory representations. La Jolla, CA: Laboratory of Comparative Human Cognition, 1981.

Levin, J. A. Proteus: An activation framework for cognitive process models. Marina del Rey, CA: Information Sciences Institute, 1976.

Newell, A., & Simon, H. A. Human problem solving. Englewood Cliffs, NJ: Prentice-Hall, 1972.

Nisbitt, R. E., & Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. Psychological Review, 1977, 84, 231-259.

Thomas, J. C., Jr. An analysis of behavior in the hobbits-orcs problem. Cognitive Psychology, 1974, 6, 257-269.

Representing Problem-Solving Episodes

Arthur M. Farley*
and
David L. McCarty**

ABSTRACT

The understanding of simple, narrative episodes in which a protagonist successfully realizes a goal through a sequence of actions is studied. In two experiments, subjects rated the acceptability of sentences of the form "The protagonist does ACT 1 in order that the protagonist could ACT 2", where ACT 2 and ACT 1 were actions from the episode. Ratings were predicted by (i.e., inversely related to) distance within a narrative representation which organizes actions into sequences (action chains) reflecting aspects of the problem-solving plan employed by the protagonist. Subjects separated action chains that had been interleaved in a text. Mishap, irrelevant, and restorative actions were not incorporated directly into an attempt structure. Corrective actions, undoing the ill effects of mishaps, were incorporated. Further research is suggested.

I INTRODUCTION

The episode has been discussed as a major, cognitive constituent of narrative discourse (e.g., Thorndyke, 1977; Rumelhart, 1975, 1977; Mandler and Johnson, 1977). An episode encompasses the situations and actions occurring during a protagonist's efforts to realize a desired goal. The goal may be that of performing a certain action or of establishing a desired situation. Table 1 presents a grammar describing a representation for successful problem-solving episodes. Rule 1 of the grammar indicates that an episode consists of a problem and its solution. Rule 2 indicates that the elements of a problem are a situational context, a triggering event giving rise to a goal, and the goal itself. The triggering event is either an action taken by the protagonist or some external occurrence in the environment (Rule 4).

Rule 6 states that the solution to the problem is represented as one or more attempts to realize the goal. Rule 7 indicates that a problem-solving attempt generates an attempt structure consisting of an action or an action preceded by one or more preactions. The goal action, that directly subordinate to an attempt, may itself be the goal or may establish the goal situation. A preaction of an action is another action which establishes one or more necessary preconditions of that action. These preconditions are implicit subgoals of the problem-solving activity. Since attention here will be focused on action interrelationships, these subgoal situations are not explicitly represented by the grammar. Finally, Rule 8 states that a preaction is an action or may be recursively expanded into an action preceded by one or more preactions of its own.

The terminal elements of our representation grammar are STATE, ACTION, and OCCURENCE. These elements are to be bound to information elements from a given episode text during understanding either by instantiation of a known, schematic episode representation or by construction of a grammar-based representation. This process corresponds to explaining protagonist's actions (Wilensky, 1978). The attempt structure generated for a given episode represents the (intentional) enablement relations existing among actions of a protagonist's problem-solving activity.

A given action of a text may be bound to several ACTION elements in the attempt structure, establishing necessary preconditions for more than one subsequent action. For example, entering a room may enable acquisition of several instruments necessary to realize the goal of an episode. Furthermore, what appears as one sequence of actions in an episode text may constitute several subsequences of actions, each establishing different necessary preconditions of a single, subsequent action. This is demonstrated in Figure 1, which presents a problem-solving episode and associated representation.

Let an action chain be a sequence of actions which is computed from an attempt structure as follows: Begin with an action bound to an ACTION element subordinate to a PREACTION element having no subordinate PREACTION elements; continue the sequence with actions bound to ACTION elements directly subordinate to successive, parent PREACTION elements; and end with the goal action. For example, the action chains of the episode presented in Figure 1 are (5,6,7,11) and (8,9,10,11). Let the distance D(a1,a2) between actions a1 and a2 of an attempt structure equal the

* Department of Computer and Information Science, University of Oregon, Eugene, Oregon; on leave at the Artificial Intelligence Center, SRI International, Menlo Park, California.

** Department of Psychology, University of Oregon, Eugene, Oregon.

(positional) index of a2 actions a1 and minus the index of a1 in a common action chain or be undefined if they are not elements of a common chain. For example, D(5,11) is 2 in the episode structure of Figure 1, while D(10,8) equals -2.

When D(a1,a2) is positive, the extent to which a1 enables a2 decreases as D(a1,a2) increases. Thus, acceptability ratings of sentences stating such enablement would be predicted to decrease. If the distance measure is less than or equal to zero or if the two actions are not elements of a common action chain, then a1 does not enable a2. Acceptability ratings should be uniformly low for sentences stating such enablement. We briefly describe two experiments addressing these predictions. A more complete discussion of the experiments and results is presented elsewhere (Farley and McCarty, 1980).

II EXPERIMENTS

Experiment 1 used two different attempt structures to generate four episode texts. The attempt structure for one-chain episodes consisted of a single chain containing seven actions. The attempt structure for two-chain episodes consisted of two action chains each containing four actions; each episode contained seven actions as the goal action was a member of both action chains. A first sentence related the triggering event and goal of the episode; a second provided setting information. The remaining sentences described the sequence of seven actions performed to achieve the goal. One sentence in the active voice was generated for each action. Two surface versions of each two-chain episode were generated. In one, the action chains were interleaved in the text; a sentence describing an action in one action chain was followed by a sentence from the other chain. In the other, the action chains were kept intact; one action chain was completely described before actions from the other were mentioned. Figure 1 is one of the intact, two-chain episodes used. Furthermore, two versions of each intact-chain text were generated, differing in which action chain occurred first.

A set of test sentences was prepared for each episode according to the following general framework: "{Action-1} in order that {pronoun} could {verb-phrase of Action-2}.", where {Action-1} and {Action-2} were bound to sentences describing actions in the episode and the {pronoun} referred to the protagonist. For one-chain episodes, the sentences differed with respect to distance in the attempt structure. All possible positive and zero distance test sentences were generated; sentences with negative distances of -1 and -2 also were generated. For the two-chain episodes, all positive distance sentences were generated such that the main-clause action preceded the purpose-clause action in one version. Sentences differed as to whether or not the two actions occurred within the same action chain.

Subjects received four episodes with test sentences and instructions; they were informed that they were to make acceptability ratings of test sentences. Marking the space next to "NA" indicated complete unacceptability (a rating of 1) while the space next to "A" indicated complete acceptability (a rating of 7). Different degrees of acceptability could be indicated by marking appropriate spaces between the two extremes. Subjects were instructed to read carefully each episode and were allowed to refer to the episode while rating sentences. Results of Experiment 1 are presented in Table 2. In short, results indicated a significant distance effect, with most pairwise comparisons yielding significant differences in predicted directions.

Within a given episode, a protagonist may perform actions which are peripheral to goal satisfaction. Two types of peripheral actions are restorative and irrelevant actions. A restorative action is one that reestablishes a normal situation in the environment disturbed by prior problem-solving. An example would be closing a purse after removing car keys or putting away a tool after using it. An irrelevant action is one that either appears to have no pragmatic utility or to be related to goals external to that of the episode. An example would be smelling a rose while mowing the lawn or turning down the oven while washing dishes. A protagonist may even perform an action that disrupts progress toward the goal of an episode and then must recover from this setback. A mishap is an action that destroys realization of a situation created by prior, goal-directed actions of an episode. Corrective actions are actions that reestablish the situation existing prior to a mishap. An example of a mishap would be dropping a tool, which would elicit the corrective actions of bending down and picking up the tool.

Peripheral actions and mishaps should not be incorporated into the attempt structure for a given episode. As such, acceptability ratings of test sentences involving such actions should be low. On the other hand, corrective actions should be incorporated as elements of a preaction subtree of (or episode subordinate to) the action enabled by the situation they serve to reestablish. As such, corrective actions do enable subsequent actions in an action chain.

Experiment 2 was designed to test the above predictions regarding these four types of actions. Two new episodes contained one restorative and one irrelevant action; a third contained one mishap and two associated corrective actions. Each episode involved only one action chain except for the second action chain produced by the two corrective actions of the third episode. A fourth episode was characterized by actions that were low in a priori associativity, with objects used in novel ways; this episode will not be discussed here. Materials and procedures were as in the first experiment.

The results indicated significant distance effects for all three stories. The Duncan test indicated the following differences among means by distance, $p < .01$: for the restorative/irrelevant stories, $1 > 2 = 3 > 4 > 5$ and $1 > 2 > 3 = 4 > 5$; for the mishap/correction story, $1 \quad 2 > 3 > 4$. These results are consistent with those of Experiment 1. Further analyses attempted to answer whether mishaps, restorative, irrelevant, and

corrective actions are incorporated into attempt structures. Two kinds of test sentences were examined for each action type. In one, the critical actions are expressed in the main clause, enabling subsequent goal-directed actions. In the other, the actions of interest occur in the purpose clause, enabled by prior, goal-directed actions. Results indicated that restorative, irrelevant, and mishap actions are not represented as being enabled by goal-directed actions; mean acceptability ratings for such sentences were consistently below 2.0. Ratings were similar for sentences stating enablement by mishap and restorative actions. Some irrelevant actions were understood to weakly enable subsequent goal-directed actions; however, mean ratings were still below 3.0. An in-depth discussion of these results, as well as those for the low-associativity episode, are presented elsewhere (Farley and McCarty, 1980).

Results indicated that subjects understand corrective actions to be part of the goal-directed behavior of the protagonist. However, the enablement relation between corrective and subsequent goal-directed actions was not as strong as the relation between the original goal-directed actions and the same, subsequent goal-directed actions. This effect may well have been produced by differing levels of a priori relational density (Graesser, 1978). In our episode, walking to a bookcase normally is more highly associated with dictionary use than is bending down (to pick up the dropped dictionary). Relational density could also account for an effect of episode in Experiment 1.

III CONCLUSION

Results of both experiments were generally consistent with predictions made in accordance with the attempt structures described by the episode grammar. The strength of enablement was inversely related to positive distances within action chains and not correlated with surface (text) distances. The action chains computed from the grammatical representations closely resemble causal event chains described by Schank and Abelson (1977). Results obtained here lend support to Schank and Abelson-like proposals, as do recent results of Graesser (1981). We demonstrate the effectiveness of an acceptability rating paradigm for assessing such structures. Our results indicate that only goal-directed actions are incorporated into attempt structures. To state that restorative, irrelevant, and mishap actions are not incorporated is not to say that they are not remembered. It merely says that such actions are not incorporated into a representation of enablement relations between actions. An attempt structure is only part of a more complex episode representation.

How else may actions differ cognitively? This question suggested an experimental paradigm in which subjects rate sentences (actions) as to their appropriateness for inclusion in an episode summary, followed by an unexpected recall task. Results of an initial pilot study were as follows: For summary inclusion, goal-directed actions are much preferred over actions not specific to the goal of an episode; within the class of goal-directed actions, instrumental actions are rated higher than movement actions, which are preferred over corrective actions. As for immediate recall, goal-directed actions yield a pattern similar to that for summarization ratings. However, mishap and irrelevant actions are recalled as well as, if not better than, goal-directed actions. The latter result suggests that such actions, though not incorporated into the attempt structure, are in a more inclusive episode representation. Such actions can be considered potentially important in determining intent of a narrative. Research confirming pilot study results is needed.

REFERENCES

1. A. M. Farley and D. L. McCarty, "Understanding Problem-Solving Episodes," Technical Report, Computer Science Department, University of Oregon, Eugene, Oregon (1980).

2. A. C. Graesser, "How to Catch a Fish: The Memory and Representation of Common Procedures," Discourse Processess, 1, pp. 72-89 (1978).

3. A. C. Graesser, S. P. Robertson, and P. A. Anderson, "Incorporating Inferences in Narrative Representations: A Study of How and Why," Cognitive Psychology, 13, pp. 1-26 (1981).

4. J. M. Mandler and N. S. Johnson, "Remembrance of Things Parsed: Story Structure and Recall," Cognitive Psychology, 9, pp. 111-151 (1977).

5. D. Rumelhart, "Notes on a Schema for Stories," in Representation and Understanding, D. Bobrow and A. Collins (Eds.). New York: Academic Press (1975).

6. D. Rumelhart, "Understanding and Summarizing Brief Stories," in Basic Processess in Reading: Perception and Comprehension, D. La Berge and S. J. Samuels (Eds.). Hillsdale, New Jersey: Lawrence Erlbaum Associates (1977).

7. R. C. Schank and R. P. Abelson, Scripts, Plans, Goals, and Understanding, Hillsdale, New Jersey: Lawrence Erlbaum Associates (1977).

8. P. Thorndyke, "Cognitive Structures in Comprehension and Memory of Narrative Discourse," Cognitive Psychology, 9, pp. 77-110 (1977).

9. R. Wilensky, "Why John Married Mary: Understanding Stories Involving Recurring Goals," Cognitive Science, 2, pp. 235-266 (1978).

Table 1

An Episode Representation Grammar

Rule 1: Episode -> Problem Solution
Rule 2: Problem -> Situation Event Goal
Rule 3: Situation -> STATE*
Rule 4: Event -> ACTION | OCCURENCE
Rule 5: Goal -> STATE | ACTION
Rule 6: Solution -> Attempt*
Rule 7: Attempt -> ACTION | (Preaction* ACTION)
Rule 8: Preaction -> ACTION | (Preaction* ACTION)

* Indicates one or more occurences of the associated element.
| Indicates alternative elements.

Bob had recently received a note from a friend (1) and wanted to send him a letter (2). He was in a post office (3) and had his friend's address on an index card in his shirt pocket (4). Bob went to a postal clerk (5). Then he bought a stamp (6). Next Bob put the stamp on the letter (7). Then he unbuttoned his shirt pocket (8). He took out the index card (9). Then he copied the address onto the envelope (10). Finally, Bob mailed the letter (11).

Table 2

Mean Acceptability Ratings

One-chain Episodes

Episode	Distance -2	-1	0
JOHN	1.45	1.26	1.55
MARY	1.11	1.20	1.69

	1	2	3	4	5	6
JOHN	6.22	5.72	4.97	4.79	4.27	4.63
MARY	6.19	5.00	4.78	3.97	4.08	4.33

Two-chain Episodes

Episode	Distance Same 1	2	3
BOB	6.48	4.99	4.65
FLO	6.43	4.81	3.75

	Different 1	2	3	4	5
BOB	1.80	1.60	1.73	1.72	2.10
FLO	1.59	--	1.96	--	1.26

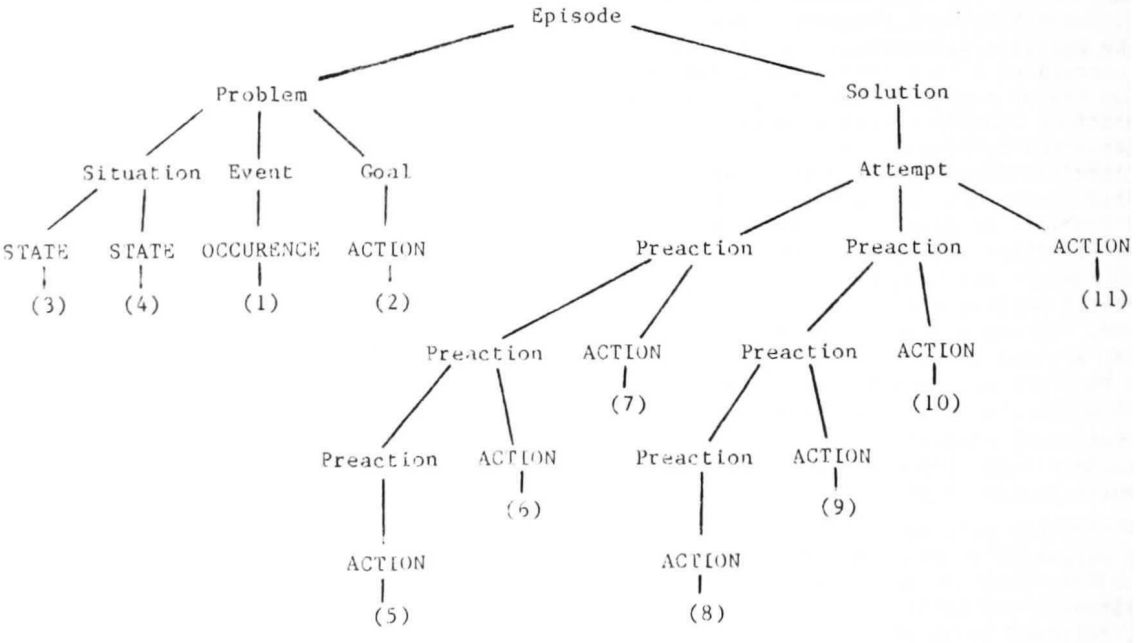


Figure 1 An Example Episode and its Representation

WHERE PROCESS- AND MEASUREMENT MODELS MEET:

Evaluation of states in problem solving

Jan Drösler, Universität Regensburg, Germany.

Summary:

Any process model of problem solving trying to capture goal directed progress needs an evaluation function of the states. Maximation of its gradient at each step serves as a decision rule in chosing among different legally possible moves. The present analysis gives sufficient conditions for the existence of an evaluation of states which is applicable to customary laboratory problems in the study of problem solving. The evaluation, moreover, is shown to establish a foundation for the measurement of "insight", the anticipatory aspect of human problem solving capacity.

Representation of states

In this study, the sets of states occuring in a problem solving task are represented by a linear module over a ring. This reduces the set of states to be discussed to those, which can be represented by vectors. In the study of problem solving this happens more often than not. Luchins' water-jar task, defines states by triples of numbers, the components of which stand for the fillings in three jars. The missionary-cannibal task represents its states as triples, containing the numbers of missionaries, of cannibals, and of the boat's disposal. For other tasks in use like the "Tower of Hanoi" numerical coding is not standard, but easily achievable. This even holds for proof problems from the propositional calculus.

Vectorial coding only makes sense if it serves a psychological purpose like permitting computation of some sort. The most simple calculation is addition. This raises the question, whether representation of states by vectors makes vector addition a meaningful operation, the result of which reflects empirical facts. At first sight it does not look that way: If in the context of the Luchins water-jar task filling is interpreted as addition, then the jars are likely to run over. Water is spilled and problem solving within the rules yet to be discussed appears to be impossible. Likewise with the missionary-cannibal task addition can lead to an uncontrolled increase of the number of missionaries or cannibals destroying the problem setting. The same appears to hold for other tasks.

Within the present study these risks are abolished by a simple device: It consists in a representation of states by finite sets. These are, e.g., with the Luchins water-jar task the integers {0, 1, 2, 3, 4, 5} or with the missionary-cannibal task the integers {0, 1, 2}. with the proof problems the integers {0, 1}. This confinement has two decisive effects for the analysis of problem solving tasks. For once in the finite case anything that can happen may be analyzed. On the other hand, an addition of states can be introduced as a meaningful representation as soon as a boundary condition is introduced. It consists in identifying the largest integer with zero. It is agreed upon, that, e.g., any three missionaries meeting waive the crossing, and that a jar filled with six units of liquid triggers a mechanism which automatically empties the jar. Under these conditions the finite set of integers assumes the structure of a residual classes ring, which is the minimal structure necessary for the introduction of vectors.

Rules as equations and inequalities

The definitions given above further the analysis of the problem solving task because the drawing rules which are usually only given verbally, now can be formulated as equations or inequalities. This possibility is never given a priori, but presupposes structural foundations like those discussed.

The proposition, calling for a constant sum of filling levels in the Luchins water-jar task only makes sense, after at least a linear module has been defined. Then filling can be represented by vector addition: Any filling is given by the vector difference of two states represented by the vectors of their triples of filling levels. Such vector difference exists for any two states, even if they are not transformable into one another by the rules. It is a system of equations, which defines the legal moves and differentiates them from the illegal ones. It can be shown, for the tasks mentioned,that this is always possible, even for the proof task from the propositional calculus, algebraization of which was introduced by Stone (1936). Addition represents filling with the water-jar task, crossing with the missionary-cannibal task and the logical exclusive "or" with the proof problems.

Because of the finite set of states, for any equation representing a drawing rule the solution set of states fulfilling this equation can be found. This set comprises single states, e.g., for one rule in the Luchins task those which have a constant sum of filling levels. It may comprise pairs of states, the second member of which can be reached from the first in one move under the rule in question.
The purpose served by algebraization of the task is two-fold: For once the discourse is no longer confined to a certain, e.g., water-jar task but rather to the unlimited set of possible tasks of this kind. The realizations, presently used in the problem solving laboratory, are then special cases with certain parameters. Another purpose, however, is reflected by the fact, that the solution sets of states fulfilling the individual rules again allow computation of useful results.

The calculus of relations

Any binary relation can be represented by a matrix over a Boolean ring. In the present case the rows (as well as the columns) stand for the finite set of states. Entries of "one" in any cell means, that the row state can be transformed into the column state in one move according to the rule under study.

Somewhat modifying a matrix calculus introduced by Copilowish (1948) permits algebraic manipulation of such matrices. Any two relations, each representing a rule, thus lead to a new relation. It represents, what is given if both rules are obeyed. The resulting relation, e.g., with the Luchins water-jar task, characterizes transformation of row state into column state in simultaneous accordance with all rules. In this connection the unitary relations first have to be transformed into binary relations by logical product, if, e.g., both, the giving and the receiving vessel have to conform to the condition of constant sum of filling levels.

Reachability of states

Analysis now turns to the question whether any column state can be reached by the given row state in a certain number of legal moves. This amounts to logical analysis whether the state can be reached in one or in two, etc. legal moves. The representation chosen here permits this analysis to be performed algebraically for arbitrarily large matrices. The result is found out for any given number of steps which, however, need not be larger than the number of states minus one. The resulting binary matrices contain entries "one" if the column state is reachable from the row state in so and so many legal steps as were analyzed in that matrix. The final step of the task analysis consists in combining these matrices into a single one the cells of which tell, in at least how many steps a row state can be transformed into the column state. This is done by pairwise comparison of successive matrices of the first kind. If the matrix of n-1 steps does not yet show transformability of the row state into the column state but the matrix in n steps at that cell carries a "one", then this n is identified as the minimal number of legal steps in question. The evaluation of each state with respect to any goal state can be read off this "reachability matrix in n steps" as the entries of the column vector associated with the goal state. This completes analysis of sufficient conditions for the existence of a state evaluation function for arbitrary goal states in a problem solving task such as they are used in the

psychological laboratory. Fig. 1 gives an example for a missionary-cannibal task with first component in the triple designating the number of missionaries, the second the number of cannibals, and the third, if 1, availability of the boat, all at the left bank. There are 2 missionaries, 2 cannibals and a boat capacity of 2. The computation of the combined drawing rule eliminates 6 of the 3·3·2 = 18 states as not reachable.

	221	210	211	200	201	110	111	020	021	010	011
210	1										
211	2	3									
200	1	2	1								
201	6	7	4	5							
110	1	2	1	2	5						
111	4	5	2	3	2	3					
020	1	2	3	2	7	2	5				
021	4	5	2	3	2	3	2	5			
010	3	4	1	2	3	2	1	4	1		
011	6	7	4	5	2	5	2	7	2	3	
000	5	6	3	4	1	4	1	6	1	2	1

Fig. 1. Symmetric reachability matrix in n steps for (2,2,2) missionary-cannibal problem computed from the rules within the module representation. Entries give minimal number of steps from row to column state.

The measurement of "insight"

For each goal, the evaluations establish a binary relation over the set of states. It can be interpreted as "... closer to the goal than ..." and is connected and transitive. It is, therefore, a weak ordering relation which plays a central rôle in measurement theory (cf. Krantz et al., 1971). The weak ordering together with some other postulates permits measurement in the technical sense of goal distance. Now, this magnitude is not in need of measurement because it can be calculated under the conditions specified. Still the cue can be taken up for the analysis of psychological data which consist of sequences of the states actually visited in the course of an experiment in problem solving. Each step can be interpreted as the result of a preference reaction over the legal steps possible at that choice point. This opens up the possibility of scaling the subjective evaluation by means of a measurement model, e.g., of ordinal measurement. Variation of the parametrization of a task creates problems of differing but controlled complexity. The problem solver's "insight" is the amount of complexity, expressed as given minimal length of solution path, the laboratory data of which do not invalidate the assumptions at the base of the measurement scaling procedure used. Empirical work by Sydow (1970) with the "Tower of Hanoi" shows, that the subjective evaluation is a function of both goal distance as well as distance from the starting point.

References:

COPILOWISH, I.M. Matrix development of the calculus of relations. J. Symb. Logic, 1948, 13, 193-203.

KRANTZ, D.H., LUCE, R.D., SUPPES, P., & TVERSKY, A. Foundations of measurement, Vol. 1. New York: Academic Press, 1971.

LUCHINS, A.S. Mechanization in Problem Solving. Psychological Monographs, 1942, 54 (whole nr. 248), 1-95.

STONE, M.H. The representation theorem for Boolean algebra. Trans. Am. Math. Soc., 1936, 40, 37-111.

SYDOW, H. Zur metrischen Erfassung von subjektiven Problemzuständen und zu deren Veränderung im Denkprozeß. Z. Psychol., 1970, 177, 145-198; 178, 1-50.

Positive Affect and Creative Problem Solving

Alice M. Isen and Gary P. Nowicki

University of Maryland Baltimore County

Abstract

Two studies run simultaneously investigated the influence of positive affect on creative problem solving as indicated by Duncker's (1945) candle task. Results show that positive affect, as induced by exposure to a funny movie, facilitated a subject's ability to solve the problem in comparison with those in control groups who either saw a control film or who did not view a film at all. In addition, in accord with previous findings (Adamson, 1952; Higgins & Chaires, 1980), subjects in another comparison group who were exposed to a facilitative display of the items were also more likely than a control group to solve the candle task. Results are discussed in terms of the influence of a positive affective state on accessibility of material and on cognitive organization.

Recently, researchers have shown a growing interest in the influence of affective states on cognitive processes. Earlier work had indicated an influence of affective states on social behavior such as helping and interpersonal attraction (e.g., Aderman, 1972; Gouaux, 1971; Isen, 1970; Isen & Levin, 1972; Veitch & Griffitt, 1976); and more recent work had suggested examining this relationship in terms of the influence of affect on the cognitive processes involved in the decision to engage in such social behavior (Levin & Isen, 1975; Isen, Shalker, Clark, & Karp, 1978). Most recently, attention has begun to focus on the influence of affect on cognitive processing more generally.

This work indicates a number of influences on cognition. For example, a state-dependent-learning effect of affective state has been observed under some circumstances (intense affect induction and maximal discriminability of the lists of words learned in the different states), as had been found earlier for other states such as drug and alcoholic intoxication (Bower, 1981; Bower, Montiero & Gilligan, 1978; Henry, Weingartner, & Murphy, 1973; Weingartner & Faillace, 1971; Weingartner, Miller, & Murphy, 1977). Work on the influence of affect on memory has also shown us that even a mild affective state of the kind likely to be experienced in everyday life (and of the kind previously shown to influence social behavior as described above) can influence cognitive processing. In addition, this influence may be even more general than that of a state-dependent learning effect. For example, a positive affective state has been shown to be capable of serving as a retrieval cue for positive material in memory, influencing such measures as the subset of words likely to be recalled from a memorized list and the reaction time for recall of affect-compatible words (Isen, Shalker, Clark, & Karp, 1978; Teasdale & Fogarty, 1979; Teasdale & Russell, 1981).

These latter findings suggest that affect influences not only memory but also cognitive organization and cognitive consequences of this organization. One study, for example, investigated the influence of affective state on judgment and evaluation. These results indicated that when people had been given a small free gift, they were more likely to judge their consumer goods more favorably; and the authors of that paper attributed these improved opinions to the affect-cued accessibility of positive material described above (Isen, et al., 1978). Positive feelings were hypothesized (and shown) to cue positive material in memory, and the presence of this material was hypothesized to influence the decision-making process with regard to affect-compatible judgments and behavior, which were shown to be affected by a positive feeling state. (This process was called a "cognitive loop," since the resulting positive judgments and behavior would then be expected to sustain the feeling state and the process.)

There is growing evidence, then, that positive affect can influence cognitive organization through what is brought to mind, and that the altered cognitive organization can then influence other ongoing cognitive processes. In the present paper, we test the hypothesis that one effect of this change in cognitive organization is to facilitate creative problem solving of the type that requires seeing things in new ways.

The task that we use is the "candle task" employed by Karl Duncker in his demonstrations (in 1945) of creative problem solving (actually, his demonstration of "functional fixedness"). In this task, the subject is presented with the common objects of a box of tacks, a candle, and a book of matches, and s/he is asked to attach the candle to the wall (a cork board) in such a way that it will burn without dripping wax on the table or floor beneath. The problem can be solved if the box is seen as an object separate from the tacks and its own properties recognized and utilized. The box can be emptied, tacked to the wall, and used as a platform for the candle, which can then be lit and will not drip wax on the table or floor.

Adamson (1952) showed that displaying the items involved in the task separately—that is, with the tacks on the table and the box empty—highlighted their independence of one another and facilitated successful solution of the problem. Recently, Higgins and Chaires (1980) found that having subjects repeat the names of common pairs of objects, but in relatively unaccustomed linguistic structures that tended to differentiate the pair members instead of in the accustomed undifferentiating structure ("tray and tomatoes" instead of "tray of tomatoes") facilitated performance on the candle task. They interpreted their results as due to the way in which the stimulus display is encoded and the increased accessibility of different constructs that could be used to characterize its elements. Thinking about the actual independence of usually-paired or united items might allow them to be utilized more completely.

We suggest that positive affect should also facilitate performance on such tasks through this same mechanism, accessibility of cognitive material. But we propose that in the case of affect, accessibility facilitates creative problem solving, not through the particular content of that which comes to mind, but through the overall amount and variety of material cued, and concomitant changes in cognitive organization and processing strategies that accompany increased accessibility of this large volume of material (Isen, 1981). We have already noted that positive affect has been found to cue positive material in memory. In addition, there is evidence that positive material tends to be more extensive and more interrelated than other material (Bousfield, 1944, 1950; Boucher & Osgood, 1969; Matlin & Stang, 1979; White, 1936; White & Powell, 1936; see also Clark & Isen, 1980). Together with the knowledge that a positive state cues positive material, this would imply that a person who is feeling good will have access to a greater amount and variety of material or ideas. This should result in the person having more ideas about how to solve a problem requiring creative inventiveness, if she is asked to solve such a problem.

In addition, if all of this is happening, a person who is feeling good will have at his disposal an increased amount of material, and it is therefore reasonable to suggest that he will organize it differently from the way he would when not so abundant with ideas. There is some evidence, for example, that decision-making style is altered by positive affect, in the direction of more efficient, or at least speedier, processing (Isen, 1981). This altered approach might well involve changes in organization of material which result in the person's seeing relationships not ordinarily seen. For both of these reasons, then, a broader array of ideas and an altered organizational structure for processing them that allows the person to relate otherwise-unrelated items, we predict that positive affect should lead to improved creative problem solving.

Method

Subjects. Subjects were 65 male and female students enrolled in classes in introductory psychology.

Procedure. Subjects were admitted to the laboratory in groups of four. They were seated and given a few minutes of introduction to the study. Then, in Study I subjects were shown a ten-minute segment of a film, either a comedy film in Condition 1 (the positive-affect condition) or a neutral film in Condition 2.

Subjects in Study II, which was run simultaneously with Study I and which thus can also be conceptualized as two additional conditions of Study I, saw no films. Instead, in this study, differing conditions were created by differences in the way in which the items presented to the subject as part of the candle task were displayed. In Condition 3, the control display was presented: a box filled with tacks, a candle, and a book of matches. In Condition 4, the same items were presented, but the tacks were displayed in a pile next to the empty box.

Following the initial phases of the study, subjects were asked to fill out a word-rating scale, in which they rated the pleasantness of unfamiliar words, as a manipulation check. Previous studies have used such indirect assessments of affect (for example, ratings of ambiguous slides), demonstrating their association with behavioral indices of positive affect such as willingness to help another person (Isen & Shalker, 1977; Forest, Clark, Mills, & Isen, 1979).

Next, subjects were seated at individual tables in the four corners of the room. The materials for the candle task were on the table, in the appropriate display, but under a cover until the task was explained by the experimenter. After reading subjects the instructions, the experimenter removed the cover from the items and gave subjects 10 minutes in which to solve the problem.

Results

Results of the manipulation check indicate that unfamiliar words were rated more positively by subjects in the positive affect condition than the control condition ($t = 2.0$; $p < .05$).

Table 1. Percent and number of Correct Solutions to the Candle Task in Each of 4 Conditions.

Comedy Film	Neutral Film	No Film Control	Facilitative Display
9/12 75%	3/15 20%	2/15 = 13%	19/23 = 83%

Table 1 presents the data showing the number and percent of subjects obtaining the solution in each condition. X^2 tests indicate that both the facilitative display condition and the positive affect condition produced significantly more solutions than did the control conditions, which did not differ from each other ($X^2 = 15.20$, $p < .01$; $X^2 = 6.09$, $p < .01$, respectively).

Discussion

Our results indicate that procedures designed to induce positive feelings can facilitate creative problem solving of the kind assessed by Duncker's (1945) candle task. Duncker spoke of people's inability to solve this task as reflecting "functional fixedness." Another way to view this task, or perhaps an elaboration of Duncker's idea, is that it involves the ability to see all aspects of the objects presented or to see among them potential relationships other than the existing ones. Our interpretation of the fact that positive affect can facilitate this process is that it does so through the increased accessibility that it affords to the large volume of material that is positive material. This results, we believe, in a greater number of ideas coming to mind when feeling good and, as a result also, an altered method of organizing and processing them that allows the person to see potential relationships not ordinarily seen.

References

Adamson, R. E. Functional fixedness as related to problem solving. Journal of Experimental Psychology, 1952, 44, 288-291.

Aderman, D. Elation, depression and helping behavior. Journal of Personality and Social Psychology, 1972, 24, 91-101.

Boucher, J., & Osgood, C. E. The pollyanna hypothesis. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 1-8.

Bousfield, W. A. An empirical study of the production of affectively toned items. Journal of General Psychology, 1944, 30, 205-215.

Bousfield, W. A. The relationship between mood and the production of affectively toned associates. The Journal of General Psychology, 1950, 42, 67-85.

Bower, G. Mood and memory. American Psychologist, 1981, 36, 129-148.

Bower, G. H., Monteiro, K. P., & Gilligan, S. G. Emotional mood as a context for learning and recall. Journal of Verbal Learning and Verbal Behavior, 1978, 17, 573-585.

Clark, M., & Isen, A. M. Toward understanding the relationship between affect and social behavior. Manuscript, 1980. To appear in A. H. Hastorf & Isen, A. M. (Eds.), Cognitive Social Psychology. N.Y.: Elsevier-North Holland, in press.

Duncker, K. On problem-solving. Psychological Monographs, 1945, 58, Whole No. 5.

Forest, D., Clark, M. S., Mills, J., & Isen, A. M. Helping as a function of feeling state and nature of the helping behavior. Motivation and Emotion, 1979, 3, 161-169.

Gouaux, C. Induced affective states and interpersonal attraction. Journal of Personality and Social Psychology, 1971, 20, 37-43.

Henry, C. M., Weingartner, H., & Murphy, D. L. Influence of affective states and psychoactive drugs on verbal learning and memory. American Journal of Psychiatry, 1973, 130, 966-971.

Higgins, E. T., & Chaires, W. M. Accessibility of interrelational constructs: Implications for stimulus encoding and creativity. Journal of Experimental Social Psychology, 1980.

Isen, A. M. Success, failure, attention and reactions to others: The warm glow of success. Journal of Personality and Social Psychology, 1970, 15, 294-301.

Isen, A. M. Positive affect, decision-making style, and risk-taking. Paper presented as part of the 17th Carnegie Symposium on Cognition: Cognition and affect, 1981.

Isen, A. M., & Levin, P. F. The effect of feeling good on helping: Cookies and kindness. Journal of Personality and Social Psychology, 1972, 21, 384-388.

Isen, A. M., & Shalker, T. E. Do you "Accentuate the positive, eliminate the negative" when you are in a good mood? Unpublished manuscript, University of Maryland Baltimore County, 1977.

Isen, A. M., Shalker, T., Clark, M., & Karp. L. Affect, accessibility of material in memory and behavior: A cognitive loop? Journal of Personality and Social Psychology, 1978, 36, 1-12.

Levin, P. F., & Isen, A. M. Something you can still get for a dime: Further studies on the effect of feeling good on helping. Sociometry, 1975, 38, 141-147.

Matlin, M., & Stang, D. The Pollyanna Principle: Selectivity in Language, Memory and Thought. Cambridge, Mass.: Schenkman, 1979.

Teasdale, J. D., & Fogarty, S. J. Differential effects of induced mood on retrieval of pleasant and unpleasant events from episodic memory. Journal of Abnormal Psychology, 1979, 88, 248-257.

Teasdale, J. D., & Russell. Differential effects of induced mood on the recall of positive, negative, and neutral words, 1981, Manuscript, Oxford University.

Veitch, R., & Griffitt, W. Good news--bad news: Affective and interpersonal effects. Journal of Applied Social Psychology, 1976, 6, 69-75.

Weingartner, H., & Faillace, L. A. Alcohol state-dependent learning in man. Journal of Nervous and Mental Disease, 1971, 153, 395-406.

Weingartner, H., Miller, H., & Murphy, D. L. Mood-state-dependent retrieval of verbal associations. Journal of Abnormal Psychology, 1977, 86, 276-284.

White, M. M. Some factors influencing the recall of pleasant and unpleasant words. American Journal of Psychology, 1936, 48, 134-139.

White, M. M., & Powell, M. The differential reaction time for pleasant and unpleasant words. American Journal of Psychology, 1936, 48, 126-133.

MEMORY IN STORY INVENTION

Natalie Dehn
Yale University

AUTHOR is a story generating program (under development) being built as a model of how human authors make up stories. Like TALE-SPIN [4], AUTHOR requires human-like knowledge of the world, but unlike TALE-SPIN, AUTHOR also requires human-like memory organization of this knowledge. The two features of human memory most essential to the AUTHOR model of story generation are (1) reconstruction, and (2) reminding. The former is responsible for the directed nature of making up stories, the latter for the author's more "fortuitous" ideas and insights.

1. The Importance of "Re"construction

Directed story invention[1] is, according to the AUTHOR model, basically a matter of

1. having some initial idea of what one is trying to invent, and

2. applying the same reconstructive memory accessing techniques used in remembering something old to develop, flesh out, and successively reformulate that idea into a complete draft. (Of course in invention one is not actually reconstructing; hence the quotes in "re"construction.)

This view of invention is, of course, basically the converse of Bartlett's theory of remembering [1]: Bartlett viewed recall as very much akin to invention, while I am suggesting that invention is very much akin to recall. My reasons for turning Bartlett's theory around, for grounding a theory of invention in a theory of recall, are twofold:

1. There currently exists a better model of recall than of invention. In particular, Kolodner has developed a working process model of reconstructive recall of episodes from long term memory as part of the CYRUS system [5]. This could serve as a basis for modeling other reconstructive memory accessing tasks, including story invention if it turns out to be one.

2. While inventiveness appears to be a widespread human capability, it does not seem to be basic or essential in the same sense that remembering, learning, and understanding are. Therefore, if the "re"constructive invention hypothesis holds, it would account for the relative cognitive luxury of inventiveness as a free byproduct of the relative cognitive neccessity of remembering.

2. The Importance of Reminding

That reminding is a natural consequence of human reconstructive memory architecture has been proposed and argued for by Schank [7]. Basically the claim is that specific memories for some input are stored in memory at the points which provide the expectations used in understanding that input, particularly those expectations that are being violated. Reminding occurs when one later processes an input that one understands in terms of a shared memory structure. Reminding is a very common phenomenon, according to this theory, though we tend to only notice its occurance when it dredges up something relatively useless.

Given a reconstructive memory architecture, anything being understood is likely to be understood in terms of a great many different structures and can be retrieved from several of them. There can thus be several ways of being reminded of any given thing, but whatever the route to the reminding, recalling the complete experience entails reconstructing the other structures used initially in understanding it.

Reminding plays an important role in laying bridges from currently active memory structures to ones usefully, though not logically, related. Reminding thus underlies and helps explains informal reasoning. In the AUTHOR model, this form of reasoning is used heavily in story generation. One can never, of course, rely on getting reminded of something useful (i.e., encountering a useful bridge), but given the structure and contents of reconstructive episodic memory, one is bound to have some useful remindings if one is doing enough memory accessing. The process of directed reconstructive invention is just such a source of memory accessing.

In the normal process of memory access stemming from reconstructive invention, an author can be reminded of (1) experiences and observations external to the story thus far made up, such as of people she has known, things that have happened to her, ways people have reacted to various situations, etc., and (2) things internal to the story thus far made up characters, props, settings, situations, reactions, etc. The former is an important source of relevant material to be incorporated into the story. The latter helps the author

1. catch problems, such as unintentional expectation violations

2. pick up development of threads she was earlier distracted from by other narrative needs, and

3. further weave existing story threads into the fabric of the evolving story beyond that deliberately intended in the top-down initial invention of those threads

---

[1] "Story invention" should be taken to mean the invention of stories and fragments thereof episodes, characters, props, settings, etc.

3. A Closer Look at Reconstruction

There is a (weak) sense in which SAM [3] was a model of reconstructive recall of stories.[2] It did not "record" stories verbatim in memory, but rather in terms of its prior schemas; in paraphrasing (recalling) such stories from these memory structures, it couldn't help but normalize/distort the stories, much as Bartlett's subjects did in reading "War of the Ghosts". By this simple model, the reconstructive process was one of filling in of details and connections not explicitly stored as part of the story itself, from the given schema. As in Bartlett's experiments, anything in the original story that was too normal was ignored (because it could always be easily inferred); anything that was too weird was remembered as weird but could not be represented well.

A better model of reconstructive memory (MOPs) evolved out of scripts, in response to Bower, Black, and Turner [2]. MOPs [7] entail decomposition during understanding, thus allowing a great deal more sharing (and hence confusions). With this model arose the additional problem of "collecting" pieces spread all over memory. Recall became largely a matter of figuring out where in all of LTM to look.

This problem was addressed by Kolodner in the CYRUS system. CYRUS has a great deal of episodic knowledge about Vance and Muskie, culled from news stories about each of them. This knowledge is "stored" in CYRUS's long term memory, from which it can be reconstructively retrieved. Consider, for instance, how CYRUS responded to the following question:

Q: Mr. Vance, has your wife ever met Mrs. Begin?
A: Yes, most recently at a state dinner in Israel.

What is especially interesting about this response is that to come up with it, CYRUS had to "deduce"[3] that a likely place for it to have "stored" such an occurance in memory, if it did so, is at some social political event, such as a state dinner. (For details of how exactly it did so, again, see [5].)

It is also interesting to consider this example of recollective reconstruction from the perspective of story invention. Suppose, for instance, an author were writing a story in which she needed to have an encounter between the wife of the American Secretary of State and the wife of the Israeli Prime Minister, or in which she needed to get these two characters at the same place at the same time how would the author set this up? Well, one plausible place to have them meet is at some social political event, such as a state dinner; more specifically, a state dinner in Israel would do nicely. Realizing that that is a likely place for the two diplomatic wives to meet (which is the hardest part of the reconstructive process) i.e., finding where to look in memory is thus the same for reconstructive recall and reconstructive invention.

4. Successive Reformulation

Reconstructive invention, like all difficult reconstruction, is a matter of **successive reformulation.**[4] When probing memory, one may, rather than finding exactly what one is looking for, come up with a partial answer plus ideas about where to look further; pursuing these ideas may again fail to immediately lead to a solution, but rather further partially specify the answer and suggest yet further ideas.

For instance, the author may have previously decided that she wants the story to be about a shy person encountering difficulties because of his shyness. This may be a good idea but it is too vague to include as is in the final story. One possible reformulation is that the shy person have to face a job as a door-to-door salesman.[5] This reformulation will itself undergo considerable reformulation. For one thing, there is still a good deal of further concretization needed - the details of his route, product line, colleagues, etc. For another, there is now a plausibility problem - why would a shy person ever become a salesman?! This plausibility question will lead to yet a further reformulation, and the "re"construction of an explanation e.g., he was forced into it, or he didn't know what he was getting into. Each of these explanations need to be further reformulated into something more concrete - the former, for instance, into the character's severe financial problems. This, in turn, may be reformulated into his being out of work and with a mother dying of cancer unless she can get some expensive therapy.

Concretization and Plausibility Maintenance are just two sources of reformulation driving story generation. Another is Dramatization (making more hang on a decision, making an action harder). Yet another is Presentation of a Narratively Neccessary Fact: As the story world and events within it develop, some facts about the storyworld will turn out to be especially causally significant, such as that a particular character is shy, or that a particular door was left unlocked.[6] In such a case, it is important to make sure that fact is introduced in the eventual story narrative, sufficiently strongly that the reader will have it available when need. One way of doing so is reformulating the fact into a complete episode (or episodes). Thus, if the critical murder scene hinges on the door being unlocked, the author will

---

[2] TALE-SPIN's invention of bears and caves was also weakly reconstructive in this same sense.

[3] "Deducing" where in memory something might be is not, of course, a matter of formal reasoning for CYRUS, but rather a matter of successive selection and application of search and instantiation strategies, as discussed in [5].

[4] This is related to the successive refinement paradigm in planning of Sacerdoti [6]

[5] This reformulation would be arrived at by reconstruction: what sort of situation might a shy person find especially stressful? How exactly a particular author would arrive at this particular reformulation (or any of the others given below) is partially a matter of her idiosyncratic memory organization and content, but the point behind all these examples is to give a flavor of the process of successive reformulation in story invention.

[6] These especially significant storyfacts are typically invented by the author as post hoc justification for something already incorporated into the storyworld. An example of this we have already seen is the invention of the cancerous mother to motivate the shy salesman. Nonetheless, they must precede what they were invented to justify, both in storyworld time and in the narrative order of presentation.

invent a secondary episode for the express purpose of introducing this fact. (This is not to say, of course, that the episode cannot also be made to accomplish other purposes.)

When something is especially important (such as a critical character trait of the protagonist), the author may want to repeat it, for emphasis. While there is a stylistic role for literal repetition, far more interesting are conceptual repetitions. Such repetitions can be produced by successive reconstruction in multiple contexts. For instance, if it is essential to the main part of the story to realize how pathologically shy the protagonist is, the author needs to communicate before that point that he is shy. She may therefore reformulate this storyworld fact into two or more episodes - for instance, the time that he crossed the street because... and the time he flunked a course because he was afraid to explain to the teacher that ....

Successive reformulation also has the interesting side effect, when viewed from the perspective of the memory accessing it entails, of greatly enhancing one's chances for dredging up useful remindings!

5. Memory and the Process of Story Invention

It should by now be apparent why human-like memory organization is needed in a model of story generation. Long term memory is, in fact, the single most important component of the AUTHOR system.

AUTHOR LTM is reconstructive and reminiscent, drawing heavily in its design both from Schank [7] and Kolodner [5]. The AUTHOR program is starting with a prebuilt version of such a memory, supplying it with prior knowledge about such things as human goals, social roles, and interactions; the prebuilt memory is also richly studded with episodic traces of (faked) experiences that would have given rise to such knowledge.

AUTHOR LTM evolves, however, in the process of story invention. This follows from a prediction stemming from the underlying Schankian theory that memory gets modified where accessed whenever something interesting results. Thus, as characters, situations, relations, etc. are invented and developed reconstructively from prior memory structures, they are remembered in appropriate places in memory. This may, in turn, lead to partial memory reorganization.

There are two important effects of such memory modifications:

1. They allow the author, in the process of making up a story, to be reminded of prior story decisions, important for reasons discussed previously.

2. They partially account for the nonduplication of stories made up by an author, without appealing to randomness: given how critical the details of memory are to the exact "reasoning" paths an author takes in story invention, this model predicts that the same person making up two stories, even if starting from the same idea, will come up with something different the second time. Human authors, of course, have their memory still further altered between stories, from external experiences.

A further prediction of this memory model concerns what is commonly referred to as "inspiration" and what is here seen as an especially useful reminding. Such a reminding experience is most likely to occur when the author has very rich indexing - which is, indeed, the case once the author has "gotten into" the story.

Yet another thing partially explained by this model is what makes a good story idea. A good idea is something that serves as an index into a rich enough part of LTM to get the reconstructive process going to the point of self-sustaining momentum; thus one person's good idea is another person's dud because of the idiosyncratic differences in memory organization and contents. There appears to be a consensus among authors (or at least among those who write about it) that good ideas are very hard to come up with deliberately, or even to recall once thought of. They are likely, rather, to be discovered fortuitously (such as in a deep and sudden feeling of insight), frequently when the author is engaged in some outside activity, and are likely to get lost again if left to their own devices. Given a good idea, though, experienced authors can sit down and start deliberately inventing.

Thus, a minor, yet critical aspect of memory for story invention, needed to supplement human reconstructive LTM is "paper memory". It is in most ways vastly inferior (memory organization and indexing being very crude) but it is just what is needed as auxiliary storage of reconstructive pointers into LTM that are themselves very hard to remember.

REFERENCES

[1] Bartlett, F. C. (1932). Remembering. Cambridge University Press, London.

[2] Bower, G. H.; Black,J.B. and Turner,T.J. (1979). Scripts in Text Comprehension and Memory, **Cognitive Psychology**, 11, 177-220.

[3] Cullingford, Richard (1978). Script Application: Computer Understanding of Newspaper Stories. Ph. D. Thesis. Research Report #116. Department of Computer Science. Yale University, New Haven, CT.

[4] Meehan, James R. (1976). The Metanovel: Writing Stories By Computer. Ph. D. Thesis. Research Report #74. Department of Computer Science. Yale University, New Haven, CT.

[5] Kolodner, Janet L. (1980). Retrieval and Organizational Strategies in Conceptual Memory; A Computer Model. Ph. D. Thesis. Research Report No. 186. Department of Computer Science. Yale University, New Haven, CT.

[6] Sacerdoti, Earl D. (1975). A Structure for Plans and Behavior. Ph. D. Thesis. SRI AI Center Technical Note 109. Menlo Park, CA.

[7] Schank, Roger C. (in press). Dynamic Memory: A Theory of Learning In Computers and People.

RECOGNIZING THEMATIC UNITS IN NARRATIVES[1]

Brian J. Reiser, Wendy G. Lehnert, and John B. Black

Yale University

Abstract

Lehnert (1980) proposed a model thematic knowledge structures called "plot units", which are structurally defined sequences of mental states, positive events, and negative events. In a clustering experiment, subjects were asked to sort 36 stories into groups. These groups were labeled by the subjects, and that data used to identify the nature of each mental category. Plot units generally provided a good fit to the clustering patterns in the data, with higher level clusters corresponding to discriminations on the nature of the outcome and judgments about the "fairness" of the protagonist.

Much research in natural language processing has utilized an event-based level of description. Schank and Abelson's (1977) model of scripts, plans, and goals is one such model, which has been both embodied in Artificial Intelligence programs (see Schank and Riesbeck, 1981 for a review) and tested in psychological experiments (Bower, Black & Turner, 1979; Gibbs & Tenney, 1980; Graesser, Gordon & Sawyer, 1979; Graesser, Woll, Kowalski & Smith, 1980). Recently, the need for a thematic level of representation has been suggested (Dyer, 1981; Lehnert, 1980; Schank, in press). In Lehnert's (1980) system, a story can be represented as a graph of overlapping "plot units". A plot unit is structurally defined, representing a unique pattern of goal interaction and goal resolution of one or more characters. Plot units are composed of three types of causally linked "affect states", representing *mental states* (states of desire), *positive events* (events that result in positive affects) and *negative events* (events that result in negative affects). Since plot units are defined as patterns of affect states, they are abstracted from specific content situations. For example, the "competition" plot unit is defined as parallel mental states between two characters (representing mutually exclusive goals) where the goal of one character is realized as a positive event, and other goal is realized as a negative event. Thus, one character winning an argument with another about which TV program to watch, and someone getting a job for which another character had applied, are both examples of the competition plot unit.

To establish initial evidence for the use of a thematic representations in processing narratives, and for the plot unit system in particular, we employed a clustering task to investigate subjects' perceptions of similarity between plots of different stories. Clustering tasks have previously been used by cognitive psychologists to study the organization of the mental lexicon (Miller, 1972), and to examine the hierarchical structure of stories (Pollard-Gott, McClosky, and Todres, 1979). In this experiment, subjects read a group of stories and were asked to sort them into groups of stories with "similar plots". Stories of varying content were constructed using the same plot units. There are many dimensions which subjects could conceivably use to judge thematic similarity: type of plan, emotions of the characters, contextual settings, personality of the main character, desirability of the story situation, etc. We expected, however, that the plot unit analysis would predict judgments of thematic similarity, and therefore that stories built of the same plot units would be grouped together by the subjects.

Method.

Thirty six two or three sentence stories were constructed from six sets of plot units. Stories constructed of the same plot units concerned different types of problems, goals, and events.

The plot units used in these stories are:

1. Stories 1-6: Competition, Denial, Retaliation, Fleeting Success. X and Y have conflicting goals, and X's are satisfied. Y asks X to agree and satisfy his goals, but the request is turned down. Y retaliates by doing something which terminates X's original success.

2. Stories 7-12: Competition, Denial, Change of Mind, Success. X and Y have conflicting goals, and X's are satisfied. Y asks X to agree and satisfy his goals, but the request is turned down. Y finds another way to successfully satisfy his goals.

3. Stories 13-18: Request Honored with Nested Promise and Reneged Promise. X makes a request and has to promise something to Y in return for Y's granting the request. However, X never performs his part of the bargain.

4. Stories 19-24: Threat, Problem Resolution. X threatens Y, who finds a way to overcome the threat.

5. Stories 25-30: Failure, Shared Negative Event. X fails in achieving some goal, and Y shares in X's failure.

6. Stories 31-36: Regrettable Mistake, Problem Resolution. X does something (accidentally) which is a negative event for both Y and X. This motivates X to do something which is positive for Y.

Thirty six Yale undergraduates were each given these stories in one of four random orders. The subjects were told to read all the stories once, and then sort them into groups, placing stories with "similar plots" into the same group. They were told to construct roughly between two and twelve groups, although they were to consider that as only a guideline. After the subjects were finished classifying the stories into groups, they were asked to label each group with a phrase that described the stories in that group.

Results.

Johnson's (1967) hierarchical clustering analysis provides a method of assessing the prototypical or average sorting of the stories, which may be used to determine whether the plot unit graphs are successful in predicting which stories will tend to be sorted together. The analysis produces a tree structure, showing the progressive merging of items empirically less and less similar (i.e., sorted into the same group less frequently) into larger and larger clusters.

The "diameter" or "minimum" clustering algorithm (Johnson, 1967) was

[1]Ray Gibbbs, Scott Robertson, Larry Hunter, and John Leddo provided valuable comments on an earlier draft of this paper. This research was supported by a grant from the Sloan Foundation.

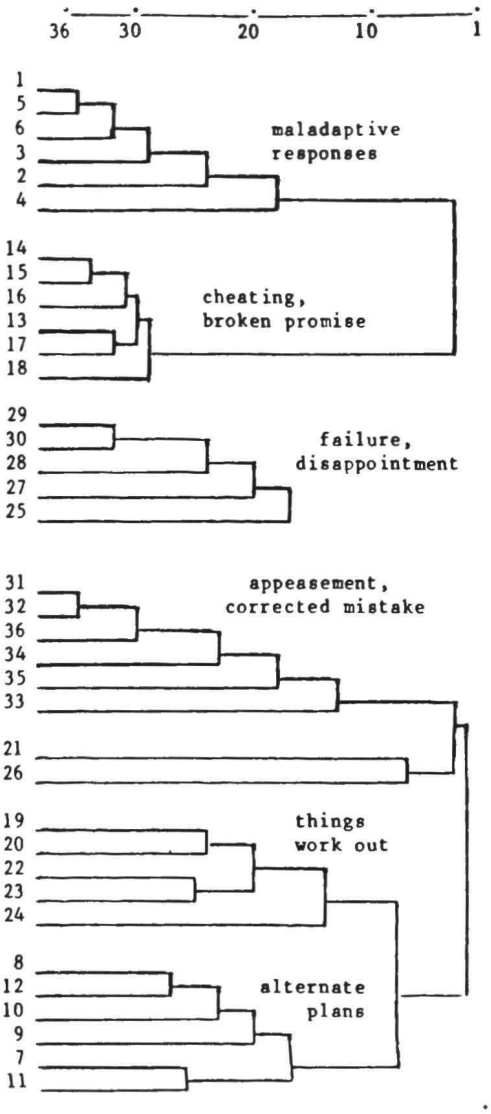


Figure 1. Hierarchical clustering diagram for the 36 stories sorted by subjects. The numbers on the horizontal axis represent similarity scores. If a cluster appears with score n, it means that each and every pair of its stories were sorted together by at least n subjects. Clusters are labeled with the most frequent tags given by the subjects that were common to all the stories in the cluster.

but leaves unspecified the nature of the problem situation (a threat). The abberant story from this group (#21) may have been sorted with the group 6 stories because the problem solution is of a more cooperative nature, making it more like correcting a mistake than reacting to a threat.

The next component of the cluster was formed of the six stories of group 2, which were labeled with "alternate plans" and either "things work out" or "rational problem solution" These labels accurately capture both the "change of mind" plot unit, and the final "success" unit, without specifying the nature of the problem.

The third component of the cluster is formed of the six stories of group 6, which were labeled with "appeasement" and "corrected a mistake". "Appeasement" indicates a simple repayment, while "corrected a mistake" more accurately reflects the regrettable mistake and problem resolution plot units.

The stories of groups 2 and 4 are more similar to each other than to the stories of group 6, although all three components are weakly clustered together. Groups 2 and 4 appear to be linked since the stories groups contain positive and reasonable goal resolutions (finding an alternative solution, overcoming a threat). Group 6 also contains a reasonable goal resolution, although it differs in that the protagonist who resolves the problem situation had inadvertently created it (for himself and another character).

Discussion.

The six clusters of stories found in the data correspond very well with the six groups of stories as predicted by the plot unit representations. Further, some of the subjects' labels accurately capture the gist of a plot unit (e.g., "broken promise", "corrected mistake"), at the same level of abstraction.

The weaker (major) clusters found in the data indicate that subjects are also sensitive to other more abstract level of representation. The three major clusters may be described by the type of outcome, and some judgment of the "justifiability" of the protagonist's actions. Thus, in the major cluster composed of groups 1 and 3, there is a malicious (and probably judged to be unjustified) action on the part of the protagonist to achieve his goal. In contrast, in the major cluster composed of groups 2, 4, and 6, the protagonist adopts a justified plan. In groups 2 and 4, the protagonist finds a rational solution to a problem, while in group 6, the protagonist adopts a positive plan to rectify a problem situation he has accidentally created. In both of these major clusters, the protagonist's goal is achieved. In the cluster of the group 5 stories, the protagonist's goal results in failure. Subjects' labels also provide support for a classification by general type of outcome ("failure", "everything works out") and judgments of fairness and motive ("maladaptive responses", "rational behavior", "malicious behavior"). Thus, in sum, the three major clusters of stories correspond to (1) the protagonist's malicious achievement of a goal, (2) justified achievement of a goal, and (3) failure to achieve a goal.

In general, it seems that subjects were more sensitive to the types of actions and types of outcomes than they were to the types of initial problem situations. Thus, descriptions of actions and outcomes were used to label the groups, (e.g., "maladaptive responses", "reneged promise", "disappointment", "failure", "things work out", "corrected mistake", and "appeasement"), but the motivation of the problem situation (e.g., the competition, threat, or denied request) not mentioned in labeling the stories. Further, the stories of groups 1 and 2, which had the same problem situation (competition, denial) but different solutions (retaliation, alternate plans) were not clustered together, while similar solutions to different problems were connected in the three major clusters.

The experiment presented here indicates that there is a thematic analysis taking place during story understanding, at least in this somewhat artificial task. It seems likely that the knowledge structures exhibited in this task would be used in a more typical understanding situation. In particular, Lehnert, Black, and Reiser (1981) have shown that plot unit representations provide a generally good prediction of subjects' summaries of narratives.

References

Bower, G. H., Black, J. B., & Turner, T. J. Scripts in comprehension and memory. *Cognitive Psychology*, 1979, **11**, 177-220.

Gibbs R. W. & Tenney, Y. J. The concept of scripts in understanding stories. *Journal of Pscholinguistic Research*, 1980, **9**, 275-284.

Graesser, A. C., Gordon, S. E., & Sawyer, J. D. Recognition memory for typical and atypical actions in scripted activities: Tests of a script pointer + tag hypothesis. *Journal of Verbal Learning and Verbal Behavior*, 1979, **18**, 319-332.

Graesser, A. C , Woll, S. B , Kowalski, D. J., & Smith, D. A. Memory for typical and atypical actions in scripted activies. *Journal of Experimental Psychology: Human Learning and Memory*, 1980, **6**, 503-515.

Lehnert, W. G. Affect units and narrative summarization. Research Report #179, Dept. of Computer Science, Yale University, 1980.

Lehnert, W. G., Black, J. B., & Reiser, B. J. Narrative summarization. Manuscript submitted to IJCAI, 1981.

Miller, G. A. English verbs of motion: A case in semantics and lexical memory. in A. W. Melton & E. Martin (Eds.), *Coding Processes in Human Memory*, Winston, 1972.

Pollard-Gott, L. P., McCloskey, M., & Todres, A. K. Subjective story structure. *Discourse Processes*, 1979, 2, 251-281.

Schank, R. C. *Dynamic Memory: A Theory of Learning in Computers and People*. in press.

Schank, R. C. & Abelson, R. P. *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum, 1980.

Schank R. C. & Riesbeck, C. K. (Eds.) *Inside computer understanding: Five programs plus miniatures*. Hillsdale, NJ: Erlbaum, 1981.

Using Qualitative Simulation to Generate Explanations

Kenneth Forbus, Albert Stevens

1 Introduction

An important goal of a computer aided instruction system is to provide students with understandable explanations. Generating explanations requires that the instructional system must itself have some understanding of the topic, prefereably close to the kind the student should have. There is a growing amount of evidence that human understanding of physical systems is based on qualitative models of those systems. This evidence comes from psychological studies [Larkin, McDermott, Simon & Simon, 1980, Stevens, Collins & Goldin, 1979] and is suppported by successes in artificial intelligence in actually constructing systems that reason about physical situations using qualitative models [deKleer, 1979a, Forbus, 1980].

Consider the following explanation of an air operated pilot valve.

> As the controlled pressure (discharge pressure from the diaphragm control valve) increases, increased pressure would be applied to the diaphragm of the direct acting control pilot. The valve stem would be pushed down and the valve in the control pilot would be opened, thus sending an increased amount of operating air pressure from the control pilot to the top of the diaphragm control valve. The increased operating air pressure acting on the diaphragm of the valve would push the stem down and - since this is an upward seating valve - this action would open the diaphragm control valve still wider. [Bureau of Naval Personnel, 1970], p.383.

This explanation is comprised of a set of events, each describing a qualitative change in some part of the device. The explanation is linearized and describes how physical effect is passed from one component to another. It ignores the true temporal changes; those things that are happening are happening continuously and simultaneously.

Explanations like the one above are an important component in teaching someone how a complex device works. This paper describes a computer system based on deKleer's incremental qualitative analysis techniques [deKleer, 1979b], that automatically generates such explanations.

2 An example explanation

Figure 1 presents an explanation generated by our system. Each panel of the explanation is drawn from the actual computer display that a student sees. Successive panels denote successive states of the display. The device described is a spring-loaded reducing valve, a common type of control device which serves to supply steam at a constant reduced pressure to a set of varying loads.

3 Incremental Qualitative Simulation

The basic idea for a qualitative simulation comes from the observation that when trying to understand or explain a device (as above), people often use a description of how parts of it change when some influence is applied to the system. The changes in physical quantities such as pressure or the position of a valve are typically described by using the sign of the derivative of the change. Thus, for a pressure, the changes are "up", "down" or "constant".

The sequence of events in such a simulation depends on how the components of the device are connected together; changes in one quantity can affect only those other quantities related to it through some sort of connection. This means that complex devices can be modelled by specifying how a set of component models are connected together. Once certain assumptions about the operation of the device are made, the effects of a change on one part can be found by local propagation through the component models of the device. This is the essence of the Incremental Qualitative (IQ) analysis formalized by deKleer for electronic circuits.

The component models we have used so far are very simple. Spaces in a device are modelled by chambers, with ports and pipes transmitting pressure changes through them. Valves are modelled in terms of changes in their openings; when the valve opening increases, the pressure in the input side decreases and the pressure in the

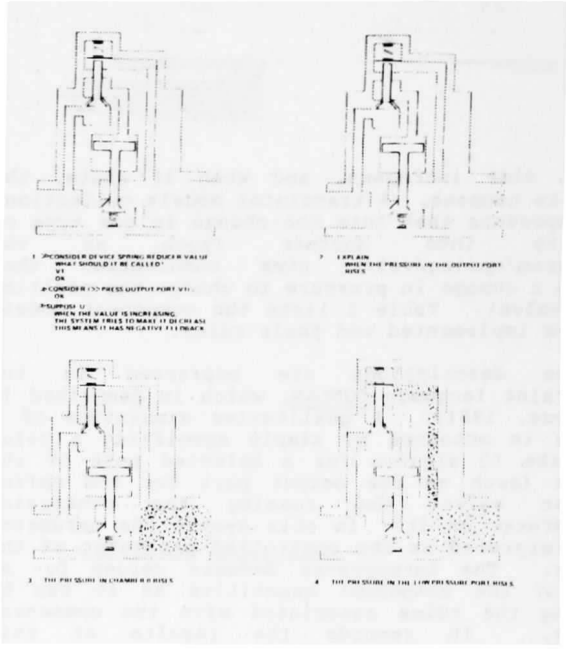
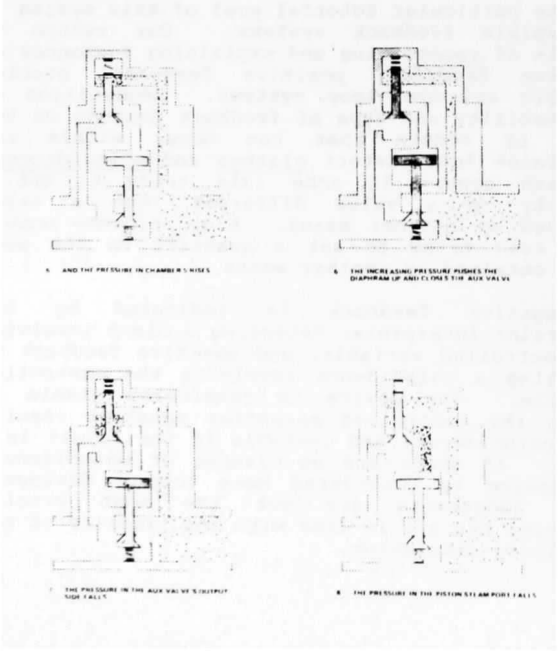
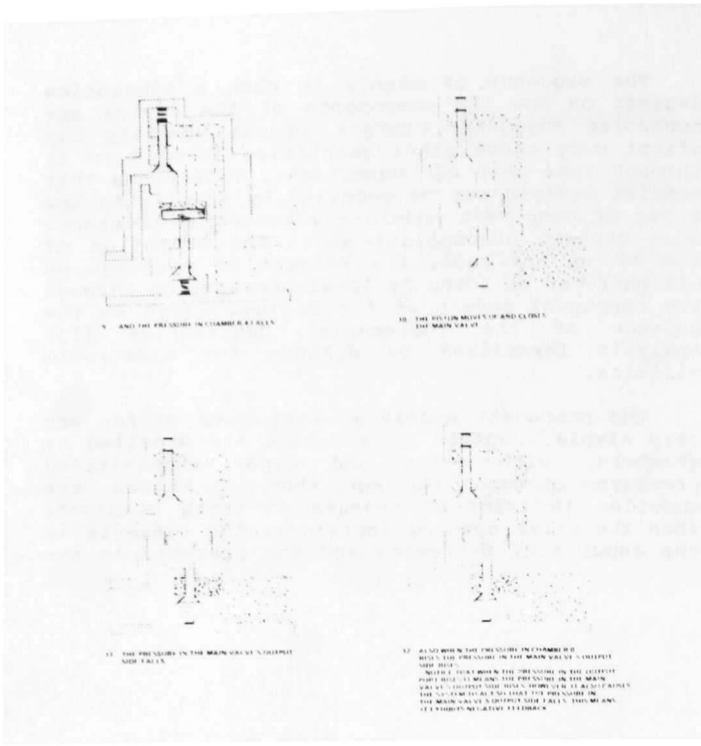


FIGURE 1 SUCCESSIVE FRAMES OF THE EXPLANATION GENERATED FOR A SPRING-LOADED REDUCING VALVE.





output side increases, and when it shuts, the opposite happens. A translator models collections of components that turn the change in one type of quantity into another (such as the diagphram/spring/valve stem combination that causes a change in pressure to change the position of a valve). Table 1 lists the component models we have implemented and their rules.

The descriptions are expressed in the constraint language CONLAN, which is described in [Forbus, 1981]. A qualitative simulation of a device is obtained by simply specifying a value from the IQ algebra for a selected part of the device (such as the output port for the spring reducer valve) and running the constraint interpreter on it. In this system the parameter is interpreted as the controlled parameter of the device. The interpreter deduces values for as many of the component quantities as it can by running the rules associated with the component models. It records the results of this qualitative simulation as a graph of the quantities, connected by the rules used to deduce them. This description of the history of the simulation is used as the basis for generating an explanation.

The particular tutorial goal of this system is to explain feedback systems. Our system is capable of recognizing and explaining instances of negative feedback, positive feedback, stable, unstable and open-loop systems. Recognition of the stability and type of feedback depends on two types of events that can occur within the constraint interpreter: clashes and coincidences. A clash occurs if some rule tries to set a quantity to a value different than a value obtained by another means. A coincidence occurs if a rule tries to set a quantity to the same value obtained by another means.

Negative feedback is indicated by the constraint interpreter detecting a clash involving the controlled variable, and positive feedback by detecting a coincidence involving the controlled variable. The device is considered stable if making the controlled parameter constant results in a coincidence, and unstable if the result is a clash. If there are no clashes or coincidences the device is considered open loop. Obviously these judgements are not the most precise possible, but are in line with the fidelity of the underlying simulation.

TABLE 1 THE COMPONENT MODELS CURRENTLY IMPLEMENTED.

The conventions are:
(1) <a> ==> <b> means
"When <a> is known, set <b> to it".
(2) <a> == <b> is equivalent to
<a> ==> <b> and <b> ==> <a>.
(3) Opposite(value) means
"If value=D then U, else
if value=U then D, else value".

One Port Chamber

Port pressure == Chamber pressure

Two Port Chamber

Portl pressure == Port2 pressure
Portl pressure == Chamber pressure

Three Port Chamber

Portl pressure == Port2 pressure
Port2 pressure == Port3 pressure
Portl pressure == Chamber pressure

Pipe

Endl pressure == End2 pressure

Continuous Valve

If valve open then opening ==>
input pressure down
and output pressure up
closing ==>
input pressure up
and output pressure down
else opening ==> valve open

(This assumes a non-zero flow)

Translator

If invert?=NO then input == output
else Opposite(input) == output

4 Generating Explanations

While the event structure of the qualitative simulation is similar to what appears to be naturally used by people, its internal form is not easy to understand. By translating it into English and using graphical cues it can be turned into a coherent explanation. This is accomplished by a simple grammar and template scheme which transforms the computation paths in the contraint network into an interleaved English and graphical presentation.

Results of analyzing the simulation are handled in the same fashion. A stored template provides an English explanation of the results, filled in with the phrases that describe the particular events in the device under consideration that led to the conclusions.

5 Conclusions

We have demonstrated that it is possible to generate coherent understandable explanations of the operation of physical devices from a qualitative simulation of the device operation. The qualitative simulation and its subsequent analysis are very general. New devices can easily be added by specifying their component connectivity and the text and graphics functions for each part.

The most important point is that these techniques make possible learning environments in which students can experiment with complex devices and see explanations of the effects of various changes. This includes changes that could not be made easily with an actual device. One could even imagine constructing a "design laboratory" that enabled students to design and experiment with a device by putting together components. This kind of learning environment could enable students to quickly understand complex physical systems in ways currently possible only after laborious study.

6 References

Bureau of Naval Personnel. Principles of Naval Engineering. : U. S. Government Printing Office 1970.

deKleer, J. Causal and Teleological Reasoning in Circuit Recognition. PhD thesis, Massachusetts Institute of Technology, September, 1979.

deKleer, J. The origin and resolution of ambiguities in causal arguments. International Joint Conference on Artificial Intelligence, August, 1979.

Forbus, K. A Study of Qualitative and Geometric Knowledge in Reasoning about Motion. Master's thesis, Massachusetts Institute of Technology, February, 1980.

Forbus, K. A Conlan Primer. Technical Note ***, Navy Personnel Research and Development Center, 1981.

Larkin, J., McDermott, J., Simon, D.P. & Simon, H.A. Expert and Novice Performance in Solving Physics Problems. Science, June 1980, 208, 1335-1342.

Stevens, A., Collins, A., & Goldin, S. Misconceptions in students' understanding. International Journal of Man-Machine Studies, 1979, 11, 145-156.

INCENTIVE AND COGNITIVE PROCESSING

Michael W. Eysenck
Birnbeck College, London

It has been found that performance on a large variety of tasks is enhanced when incentives (monetary or otherwise) are offered for high performance efficiency. Several possible theoretical explanations have been proposed to account for this effect. Edwin Locke has argued that incentives improve performance to the extent that they affect the individual's goal-setting; in contrast, Easterbrook claimed that incentives produced increased attentional selectivity with enhanced performance on the primary task being accompanied by reduced performance on a concurrent secondary or subsidiary task.

An influential theoretical position deriving from that of Easterbrook was put forward 10 years ago by Donald Broadbent. He pointed out that incentive often interacts with known arousers (e.g., white noise) in such a way as to suggest that incentive is itself an arouser. The implication is that various arousers, including incentive, are affecting the same arousal mechanism in very much the same way. In this connection, there is some evidence indicating that incentivized subjects are more physiologically aroused than control subjects.

One of the major limitations of the research in this area has been the failure to assess the effects of incentive on performance efficiency in a satisfactory manner. If incentive improves performance on a cognitive task, it does not necessarily follow that incentive has enhanced the efficiency with which all of the component processes involved in the performance of that task have been carried out. Indeed, it is even possible for an overall beneficial effect of incentive on the performance of a cognitive to mask an adverse effect of incentive on one or more of the component processes.

The cognitive task used in our ongoing research program was selected in part because it permits identification and measurement of its salient component processes. It was also selected because the effects of one arouser (white noise) on its performance have already been established by Hamilton and Hockey, thus permitting some evaluation of Broadbent's arousal-based interpretation of incentive effects. The task involves letter transformation; more specifically, the subject is given one, two, three, or four letters, and is asked to add either 2 or 4 letters to each of the original letters. Thus an example of a simple problem is 'C + 2', for which the correct answer is 'E', and an example of a relatively difficult problem is 'JEPG', for which the correct answer is 'NITK'. For each problem, the subject must work out the entire answer before responding.

What are the component processes involved in this letter-transformation task? When a letter is presented, the first process involves accessing long-term memory, locating the alphabet, and then finding the appropriate starting point within the alphabet. The second stage of processing involves the carrying out of the transformation itself, and the third stage of processing involves the storage and organization of the part-answer. In the case of a four-letter problem, these three processing stages are repeated for each letter in turn. Thus, at least conceptually, we can sub-divide the total time taken to solve a four-letter problem into 12 component stages.

A further advantage of this cognitive task is worth mentioning at this point. While it is often extremely difficult (or even impossible) to decide whether different task-processing activities occur serially or in parallel, it is virtually certain that

the component processes involved in the letter-transformation task are carried out seriatim. It is hard to see how the transformation stage could begin before the alphabet has been located in long-term memory, and it is equally difficult to believe that the answer to a letter could be stored while it is still being transformed.

In the initial study in the current series, subjects spent 5 minutes solving each of 8 different versions of the task (1, 2, 3, or 4 letters, adding 2 or 4). Incentive was a between-subjects' factor, with incentivized subjects being offered £5 (approximately 12 dollars) for obtaining an overall level of performance among the top 25% of participating subjects. Non-incentivized subjects were offered no extra monetary payment over and above their normal payment for experimental participation.

The results of this initial experiment were reasonably unequivocal. The incentivized subjects outperformed the non-incentivized subjects in each of the 8 task conditions, taking between 30% and 40% less time to solve each problem (the error rate was less than 5% in all conditions). The only significant interaction was between incentive conditions and the number of letters in each problem; this interaction involved a systematic increase in the beneficial effect of incentive as the number of letters requiring processing increased.

What do these results mean? In order to interpret the interaction between incentive conditions and number of letters per problem, we obviously need to have some understanding of the effects on the processing system of varying the number of letters. Perhaps the major effect of increasing the number of letters in the task is to increase the demands on some short-term storage system which is involved in the storage and organization of the accumulating part-answer. If so, then it may tentatively be concluded that monetary incentive increases the efficiency of a short-term storage system.

Of course, this cannot be the whole answer. Presumably one-letter problems make minimal demands on short-term storage, and yet incentive increased performance speed considerably on such problems. The implication is that incentive also affects time to access long-term memory or transformation speed (or both).

The second experiment in the series was designed to clarify the precise effects of incentive on the letter-transformation task. Only four-letter problems were used (adding 2 or 4), and the presentation was on a letter-by-letter basis. The following sequence of events occurred on each trial: the subject pressed a key in order to present the first letter; he or she then did the transformation aloud; then, the subject pressed a key in order to present the second letter; and so on. Each subject spent 40 minutes doing the task (20 minutes on each version). Incentive was manipulated as a between-subjects' factor, with incentivized subjects being offered £5 (about 12 dollars) for obtaining an overall level of performance among the top 25% of the subjects. Non-incentivized subjects only received their normal payment for attending the experiment.

The first step in the analysis of the data was to calculate 12 intra-task times. This was done by measuring the time between the first key press and the start of the transformation (assumed to reflect access time to long-term storage), the time to perform the transformation out loud (transformation speed), and the time between the end of the transformation and the next key press (assumed to reflect storage and organization). These three times were obtained for each of the four letters. The error rate was again below 5 per cent.

The major findings were quite straightforward. Incentive did not affect the time taken to access long-term storage, but did lead to increased transformation speed, especially when the add factor was 4 rather than 2. In

addition, incentive speeded up the time taken to perform storage and organization operations. This main effect was qualified by a significant interaction between incentive and letter position. In this interaction, the beneficial effects of monetary incentive were greatest during storage and organization following transformation of the third letter. In general terms, the demands on short-term storage capacity are likely to increase systematically with each additional letter. However, there is a reduced requirement for storage and organization following transformation of the fourth letter, since at that point the subject is in a position to output his or her answer to the problem. Accordingly, the interaction between incentive and letter position may be interpreted as reflecting the greater efficiency of some short-term storage system under incentive conditions.

It is interesting to compare the effects of monetary incentive and white noise on this task. Hamilton and Hockey found that noise increased the speed of transformation but decreased the speed of storage and organization. While more levels of noise and incentive must be sampled before any definite conclusions are possible, it is nevertheless interesting to note the rather different patterning of the effects of incentive and noise. In particular, incentive <u>increases</u> the speed of storage and organization, whereas noise <u>decreases</u> it. It is thus possible that noise and incentive should not be considered merely as equivalent arousing agents.

In the third experiment in the series we looked at the effects of distraction on the performance of the 4-letter, add-4 version of the letter-transformation task. The task was carried out in the presence of auditorily presented distracting stimuli (letters, numbers, or meaningless blips) which were presented on average one every 5 seconds, or in the absence of distraction. There were three within-subjects' incentive conditions: no incentive; £9 (about 20 dollars) distributed among the top 25% of subjects on low-incentive trials; and £70 (about 160 dollars) distributed among the top 25% of subjects on high-incentive trials. The session lasted approximately one hour.

One of the reasons for investigating the effects of distraction was that Easterbrook argued that incentive leads to increased concentration on task-relevant stimuli, which seems to imply that incentive should reduce distractibility. An alternative possibility is that more of the available processing resources are invested in the task under incentive conditions. If the active rejection of intermittent distracting stimuli requires processing resources, then incentivized subjects might be <u>more</u> rather than <u>less</u> distractible. A further possibility is that incentive interacts with type of distraction, so that incentive can either increase or decrease distractibility depending on the nature of the distracting stimuli. The data from the third experiment, which are currently being analyzed, will provide answers to some of these issues.

In sum, it is erroneous to assume that incentive produces an across-the-board improvement in all of the processing operations involved in the letter-transformation task. What actually happens is that simple mental operations such as those involved in transformation are speeded up by incentive, and the efficiency of some short-term store is improved. However, another processing operation (accessing long-term memory) is unaffected by incentive, perhaps because it is a relatively automatic skill. It is only by doing fine-grain analyses that one can obtain important information about the precise patterns of effects produced by incentive. The above findings have been obtained with the use of relatively modest incentives, of course. We have preliminary data suggesting that larger incentives may produce either somewhat different or very different results (as would obviously occur if one hundred thousand dollars were offered for good performance).

The Role of TAUs in Narratives

Michael G. Dyer
Computer Science Department
Yale University, New Haven CT[1]

1. Introduction

People often rely upon common sayings, or adages, when asked to characterize stories (either by way of summarization, or title selection). What are people doing in such cases? Why do adages often serve as an effective way of characterizing a story, and how are people able to accomplish this?

For instance, when asked to characterize the following story:

MINISTER'S COMPLAINT

> In a lengthy interview, Reverend X severely criticized President Carter for having "denigrated the office of president" and "legitimized pornography" by agreeing to be interviewed in Playboy magazine. The interview with Reverend X appeared in Penthouse magazine.

readers often responded with adages such as:

ADG-1: The pot calling the kettle black.
ADG-2: Throwing stones when you live in a glass house.

Clearly, these adages **are** an effective characterization of MINISTER'S COMPLAINT. But how do we recognize this fact? By what process does an 'appropriate' adage come to mind, and to what purpose?

Furthermore, when supplied with an adage and a context, some individuals experience remindings from episodes in their lives. For instance, one individual was first presented with the following:

context: EDUCATION

ADG-3: Closing the barn door after the horse has escaped.

and then asked to recall some episode from his life. He experienced this reminding:

ACADEMIA

> Years ago, I was at University U-1, where I could never get the facilities I needed for the research I wanted to do. So I decided to apply to University U-2, which offered a much better research environment. When the chairman learned I had been accepted to U-2 and was actually leaving U-1, he offered to acquire the facilities I had wanted. By then, however, my mind was already made up.

Several observations are worth making here: First, for adage ADG-3 to have initiated this reminding, the ACADEMIA episode must have somehow been indexed in long-term memory **in terms of some abstract situation** characterized by that adage. Furthermore, this indexing could **not** have had anything to do with the specific semantic **content** of the adage, since ADG-3 ostensibly concerns a farmer, a horse and a barn door. In contrast, ACADEMIA involves a chairman, a researcher, and university facilities.

To account for such phenomena, I will present a class of knowledge constructs, called **TAUs** (Thematic Affect Units), which share similarities with other representational systems under development at Yale, such as Schank's TOPs [8] and Lehnert's Plot Units [4] [5].

2. Thematic Affect Units

TAUs were first developed in the context of BORIS [3] [2], a computer program designed to read and answer questions about narratives that require the application and interaction of many different types of knowledge. In BORIS, **TAUs** serve a number of purposes: First, they allow BORIS to represent situations which are more abstract than those captured by scripts, plans, and goals as discussed in [7]. Second, **TAUs** contain processing knowledge useful in dealing with the kinds of planning and expectation failures that characters often experience in narratives. Finally, **TAUs** also serve as episodic memory structures, since they organize events which involve similar kinds of planning failures. For more detail on the use of **TAUs** in narratives, see [1].

In general, **TAUs** arise when expectation failures occur due to errors in planning. As such, they contain an abstracted planning structure, which represents situation-outcome patterns in terms of: (1) the plan used, (2) its intended effect, (3) why it failed, and (4) how to avoid (or recover) from that type of failure in the future. If we abstract out this planning structure from both the BARN-DOOR and ACADEMIA episodes, we get the following **TAU**:

TAU-POST-HOC

(1) x has preservation goal G [7] active since enablement condition C unsatisfied
(2) x knows a plan P that will keep G from failing by satisfying C.
(3) x does not execute P and G fails.
x attempts to recover from the failure of G by executing P.
P fails since P is effective for C, but not in recovering from G's failure.
(4) In the future, x must execute P when G is active and C is not satisfied.

TAU-POST-HOC captures the kind of planning failure that occurred for both the farmer who lost his horse, and the chairman who lost a graduate student. If the ACADEMIA story were told to an actual farmer who had lost his horse under the same planning circumstances, that farmer might well be reminded of his own experience. Whether this occurs or not, however, depends upon what other episodes are in long-term memory and what features are shared between them. Notice, for instance, that both BARN-DOOR and ACADEMIA share goals at some level. That is, both the farmer and the chairman had a goal requiring proximity on the part of another entity. Since these features are shared, one experience has a better chance of causing a reminding of the other to occur. For instance, the farmer would have recalled the HIRED HAND episode below before recalling the BARN-DOOR episode because of their shared features:

HIRED HAND

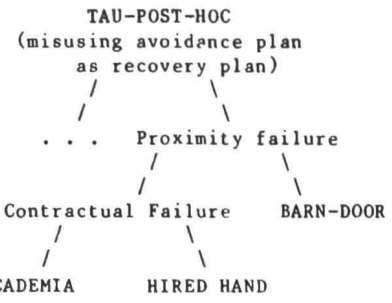
> The hired hand always wanted a raise, but the farmer would not grant it. Finally, the hired hand got an offer to work at a neighbor's farm. When the farmer found out, he offered the hired hand a nice raise, but it was too late.

---

[1]This work supported in part by the Advanced Research Projects Agency under contract N0014-75-C-111 and in part by the National Science Foundation under contract IST7918463.

Thanks go to Tom Wolf and Marty Korsin for helping with some of the ideas presented here, and for being sources of remindings.

Although these episodes (i.e. HIRED HAND, ACADEMIA, BARN-DOOR) share the same **TAU**, HIRED HAND and ACADEMIA have more indices in common. One possible organization for them appears below:



In this way **TAUs** can account for cross-contextual remindings (as in the case of BARN-DOOR and ACADEMIA). Episodes are often related in memory because they share the same abstract planning error even though they differ in content. However, cross-contextual remindings can occur only where episodes are organized under the same **TAU**, yet do not share content features. Where content is shared, the "closer" episode will be recalled.[2] Consider the following episode:

IRANIAN EMBASSY

> While holding 52 US hostages in Iran, the Iranian government condemned the take-over, by terrorists, of its embassy in Great Britain. "This is a violation of international law", protested Iran.

A reader was spontaneously reminded of this episode while reading MINISTER'S COMPLAINT (on page 1). Again, there is little in common between these stories at the content level. IRANIAN EMBASSY is about politics, while MINISTER'S COMPAINT is about pornography. However, at the abstract planning level, they both share the following **TAU**:

TAU-HYPOCRISY

x is counter-planning against y
x is trying to get a higher authority z to
  either block y's use of a plan P-1
  (or to punish y for having used P-1)
  by claiming that P-1 is an unethical plan
y claims that x has used an unethical plan P-2
  similar to P-1
therefore, x's strategy fails

In the case of MINISTER'S COMPLAINT, x is Reverend R, y is President Carter, and the third party is 'public opinion'. In the case of IRANIAN EMBASSY, x is the Iranian militants, y is the British terrorists, and the third party is 'world opinion', such as the United Nations.

As argued in [8], the reminding process is useful for this reason: Once a situation has caused one to be reminded of an episode, all of the expectations associated with that episode become available for use in making predictions about what will occur next. In the case of **TAUs**, their associated expectations include advice on either how to avoid making the error predicted by the **TAU**, or on what alternative plan can be used to recover from the error once it has been made. The ability to store **cross-contextual** episodes make **TAUs** very general and powerful mechanisms. Once an episode has been indexed under a **TAU**, its recovery/avoidance heuristics become available for use in completely different situations. Thus, planning advice learned in one context can help processing in other contexts, if the experience was recognized in terms of an appropriate **TAU** in the first place.

3. Bad Planning is Widespread

An examination of adages reveals that many are concerned with planning failures. That is, adages advice us either how to recover from a failure, or how to recognize and thus avoid future failures. Often, this advice is given implicitly, simply by describing situations in which certain planning errors lead to goal failures. In most cases, adages capture what has been called **meta-planning** [10] -- i.e. planning advice on how to select or use plans in general. For example, some adages deal with the need for checking enablement conditions before plan execution:

ADG-4 Don't count your chickens before they're hatched.

Other adages stress choosing less costly avoidance plans over more costly recovery plans:

ADG-5 A stitch in time saves nine.

or weighing the risks involved with the goal to be achieved:

ADG-6 If it ain't broke, don't fix it.
ADG-7 The cure can be worse than the disease.

Many plans require cooperation or coordination with others. This can simplify planning but complicate plan execution:

ADG-8 Two heads are better than one.
ADG-9 Two many cooks spoil the broth.

Some plans involve selecting an appropriate agent:

ADG-10 The blind leading the blind.
ADG-11 Who pays the piper calls the tune.

Timing, enablement conditions, cost, plan coordinations, and agents are just a few of the areas in which plans can go wrong. Other areas, for example, include counter-planning against a foe,

ADG-12 Cut off your nose to spite your face.

anticipating planning failures when using high risk plans,

ADG-13 Don't burn bridges behind you.

the timing of plans,

ADG-14 The early bird catches the worm.

and tradeoffs between short-term and long-term planning strategies:

ADG-15 If you can't lick 'em join 'em.
ADG-16 Don't bite the hand that feeds you.
ADG-17 Honesty is the best policy.
ADG-18 Live by the sword, die by the sword.

Any story that involves these kinds of planning failures will end up being indexed under a **TAU** which contains abstract planning advice (and can be expressed in natural language by an adage.) When a related story is read and indexed under that **TAU**, its associated adage may come to mind. For instance, a story about how a ghetto riot protesting bad economic conditions resulted in black businesses being burned, would be indexed under TAU-GREATER-HARM, with an adage such as ADG-12 possibly coming to mind.

Plans and plan failures cut across all knowledge domains. This is because we are always choosing plans, adjusting old plans to new situations, recovering from errors in planning, finding explanations for why a plan failed, etc. Furthermore, we have a large storehouse of heuristic plans, and

---

[2] This does **not** imply that BARN-DOOR can't remind one of episodes unrelated to TAU-POST-HOC. Clearly, other indexing methods may be operating at the same time. The farmer may recall BARN-DOOR simply in terms of "experiences I've had with horses". Of course, this kind of indexing can not lead to cross-contextual remindings.

there are many ways a plan can go wrong: You can't execute one plan until you have the right enablements satisfied; plan components must be executed in the right order; plans require agents, etc. This large and complex domain serves as a perfect terrain in which to index many episodes.

Many of these adages give what may appear to be superficial advice. It may seem strange that memories should be organized around such 'obvious' rules for planning, but then again, how often do we fail in our plans because we have violated some adage? How often, for instance, have we failed because we acted before we planned? ("Look before you leap.") How many times have we gotten into trouble for being late? When have we initiated a plan, only to discover we had miscalculated the amount of effort (or the side-effects) involved? ("Easier said than done.") How often have we delayed executing a simple plan, only later having to execute a more costly plan? The answer is: "very often". These adages are common because they point out the kinds of planning errors **people are always making.** By definition, plans which failed were "bad" plans. Good planners at the very least follow the general planning advice represented in the adages of their culture.

4. TAU Implementation

The recognition of **TAUs** is complex. Clearly, goals and plans must be tracked. In many cases there is also an affect component. For instance, in TAU-POST-HOC it is the futility of the recovery plan, combined with the sense of "if only I had done things differently" that helps provide an access "key" to this **TAU.**

So far the BORIS project has emphasized the use of **TAUs** in narrative comprehension. Much work remains to be done in modeling remindings during comprehension. This is important for extracting the 'moral' or point of a story. A computer program which can only answer questions of fact about IRANIAN EMBASSY, such as:

Q: How many Americans are being held in Iran?
A: Fifty-two.

Q: Who seized the Iranian embassy in Britain?
A: Terrorists.

Q: What did the Iranians do?
A: They protested the take-over.

is missing the **point** of why the IRANIAN EMBASSY is of interest. The point of IRANIAN EMBASSY is TAU-HYPOCRISY, and **that's** where it should be remembered in long-term memory (rather than just under "things I know about Iran", or "embassy events I have read").

5. TAU Experiments

What is the psychological validity of **TAUs**? Do people have **TAUs** "in their heads" and, if so, how do they use them? Some initial exploratory experiments by Seifert [9] in the Yale psychology department indicate that people use **TAUs** to organize narratives.

In one experiment, subjects read groups of stories each sharing the same **TAU**, but differing in content. Subjects were able to generate new stories, using completely different contexts, yet capturing the same planning structure specified by each **TAU.** In a follow-up experiment, different subjects were asked to sort the resulting stories generated from the first experiment. A cluster analysis [6] revealed a strong tendency for subjects to sort stories together by **TAUs.** Where stories shared the same content (but not the same **TAUs**) they were still grouped by **TAUs.**

6. Conclusions

In this paper I have presented a class of knowledge constructs, called **TAUs**, which are related to TOPs [8] and PLOT UNITs [5]. I have argued that **TAUs** organize episodes around failures in planning, and as such, **TAUs** account for at least one form of cross-contextual reminding phenomena. Furthermore, **TAUs** have adages associated with them, which express avoidance and/or recovery advice available once the **TAU** has been accessed. Since stories are indexed in terms of planning errors, this information often captures the moral or point of a story.

REFERENCES

[1] Dyer, Michael G. Thematic Affect Units and Their Use in Narratives. paper submitted to IJCAI-81, 1981.

[2] Dyer, Michael G. **In-Depth Understanding: A Computer Model of Memory for Narrative Comprehension.** PhD Thesis, Computer Science Department, Yale University, (forthcoming).

[3] Lehnert, Wendy G., Dyer, Michael G., Johnson Peter N., Yang, C. J., and Steve Harley. BORIS -- An Experiment in In-Depth Understanding of Narratives. Technical Report 188, Yale University. Dept. of Computer Science, 1980.

[4] Lehnert, W. G. Affect Units and Narrative Summarization. Technical Report 179, Yale University. Dept. of Computer Science, 1980.

[5] Lehnert, Wendy G. Plot Units and Narrative Summarization. Cognitive Science, in press.

[6] Reiser, Brian J., Lehnert, Wendy G., and Black, John B. Plot Units and the Understanding of Narratives. Cognitive Science Technical Report, Yale University (in preparation).

[7] Schank, Roger C. and Abelson, Robert. Scripts, Plans, Goals, and Understanding. Lawrence Earlbaum Associates, Hillsdale, New Jersey, 1977. The Artificial Intelligence Series.

[8] Schank, Roger C. Language and Memory. Cognitive Science 4(3), July, 1980.

[9] Seifert, Colleen. Preliminary Experiments on TAUs. unpublished manuscript. Psychology Dept. Yale University, 1981.

[10] Wilensky, Robert. Meta-Planning: Representing and Using Knowledge About Planning in Problem Solving and Natural Language Understanding. Technical Report Memo. No. UCB/ERL M80/33, Electronics Research Lab. Engineering College University of California, Berkley, 1980.

CONTROLLING PARSING BY PASSING MESSAGES

Brian Phillips & James Hendler
Texas Instruments, Inc.

1.0 INTRODUCTION

The functional segmentation of linguistic knowledge, into rules about form and rules about meaning, has been vital in unravelling the complexities of language. However, it does not follow that the process of analysis will respect the same boundaries and so the very segmentation that provided the insights can be troublesome when one seeks to create a dynamic model of language. Here we are concerned with how linguistic knowledge is used rather than what knowledge is used. The study of process is one of the contributions of a computational linguistics to the study of language (Hays, 1971).

Improper control strategies result in one of the apparent paradoxes of current knowledge-based systems: the more knowledgeable they are, the more inefficient they become. The systems are unable to handle the combinatorial explosion of possibilities that searches of the knowledge space produce. This situation is obviously counter to intuition, and to human behavior: more knowledgeable systems should perform better.

We believe that a language understanding system should have the ability to bring syntactic and semantic knowledge to bear on the analysis at many points in the computation. This enables it to resolve the alternatives as soon as possible and prevent the flow of extraneous analyses to later phases.

Existing conceptual analyzers fall into three major categories: linear control parsers (Woods & Kaplan, 1971), semantic grammars (Hendrix, 1977), and semantic analyzers (Schank, 1975). None permits the flexible, data-governed interaction of syntax and semantics.

Our approach to creating such a model is to use the notion of a society of communicating, knowledge-based, problem-solving experts, called "actors" (Hewitt, 1976). These actors can communicate by passing messages to any other actor in the system. This flexible control structure allows actors at any level of the analysis to interact with actors at other levels. The routing of information in the system is determined by the content of the linguistic act, thus achieving a data-driven control structure. The ability to direct information to the actor that needs it, is expected to improve efficiency. Each subsystem in language contributes to the process of understanding, but often offers several different views of the data. In our scheme, actors communicate to achieve a mutually consistent analysis, out of which comes an understanding of the data.

The capabilities of each actor are determined by the functional segmentation, e.g., an actor can have the ability to use constituent rules that describe the structure of a noun phrase, thus retaining this important facet of linguistic theory.

2.0 DESIGN FEATURES

In the system we are trying to achieve several other design goals in addition to the gain in efficiency:

We agree with Schank (1975) that the goal of analysis is not to produce a parse tree. It should not even be a subgoal, as is the case in systems that first produce a parse tree, then perform semantic interpretation. The parse tree should be considered a data structure that is constructed incidentally to the analysis, or can be constructed if it is needed. But syntax cannot be ignored; it is often useful in determining antecedents of proforms, for example.

Schank's (1975) hypothesis of semantic prediction appears to be a good approach. But one cannot always have expectations. We envisage a system that can flow into a predictive mode when the situation is appropriate, with a default control structure of syntax-then-semantics.

The output of the system should be semantic description of the input as instantiated case-frames. The novelty of the situation is captured by the way in which these case-frames are linked and by their spatio-temporal settings. The semantic description augments the encyclopedia, the store of world knowledge, and is thus available as pragmatic knowledge in the continuing analysis of the input.

3.0 THE ACTORS

Most of our actors are "experts" on aspects of the primitive organizing principles of syntax and of semantics. They become associated with domain knowledge, i.e., the grammar of a language, or world knowledge for a problem area. The job of an actor is to instantiate a model it has been given (top-down analysis), or if it was not given a model, then to find a model (bottom-up analysis). The process of instantiation is performed by eliciting information from other actors that can use their expertise on the problem; they, of course, may have to consult still other actors.

3.1 The Syntactic Experts

The organizing principle of syntax is constituency; the principal actor in syntax thus uses the constituency rules of the grammar to associate words into higher level constructs. The constituent actor recognizes syntactic constructions primarily by matching words to syntactic rules using the dictionary entries of words to determine their syntactic categories.

3.2 The Semantic Experts

The primitive organizing principles of conceptual knowledge are relations of sequence, contingency, enablement, equivalence, taxonomy, part-whole, etc. (Phil[illegible] how [illegible]

expert, for example, knows how to use "contingency", "sequence", and "enable" links.

3.3 Translation Experts

The actors have vocabularies that are peculiar to their domains. Therefore, messages may require translation from the terminology of the sender to that of the receiver. Take, for example, messages between clause actors (CLA) and case-frame actors (CFA). The former uses concepts like subject, object, and verb, whereas the latter uses event, state, agent, etc. There are special actors in the system to handle this.

4.0 A FRAGMENT OF AN ANALYSIS

We will show how the system analyzes:

(1) The left front tire is flat.
(2) I will change it.

and determines the referent of "it" in (2).

The goal of the system is to create a meaning representation by instantiating a CF. Through equivalency and part-whole relations, a CF can be equivalent to a complex of CF's; thus the top-level instantiation may be achieved by instantiating the lower rank CF's.

A CFA normally has a model of a CF that it is trying to instantiate. Initially this cannot be the case and the system has to revert to a bottom-up approach. The CFA sends a message to the CLA requesting that it be sent a translation of a syntactic analysis of a clause. The CLA has to find a clause using the rules of the grammar in. A series of instances of the constituent actor are invoked to analyze the rules. As they process the rules, they simultaneously notify an "input actor" of the terminal categories that they have encountered.

When all necessary constituents have been expanded, the constituent actors are halted, waiting to know which of the parse paths might be consistent with the input. The input actor prompts the user for a word. It then sends messages to constituent actors to cause the deletion of paths inconsistent with the input and then messages to those constituent actors that still have valid paths. Effectively there is parallel processing synchronized by the input.

When a clause, i.e., (1), has been found, a translation can be sent to the case-frame expert, but first it must be translated as discussed in the previous section. A request is sent to the translation expert from the clause expert. This translation can be sent to the case-frame expert directly.

The CFA next knows to ask Chronology for the NEXT-EVENT. Chronology predicts that "change tire" will be the next act. The Chronology Expert now passes this information back to the CFA.

The CFA has now processed the first case frame to the best of its abilities and sets out to instantiate the prediction. It is now working in a top-down manner. When the prediction is passed to the CLA and translated, "tire" will be available as a match for the pronoun "it".

5.0 OTHER TOPICS

It is our belief that our message-passing control structure can yield more than improved efficiency. Several of the more difficult problems in natural language processing come about due to the inability of different types of knowledge to be brought to bear at the same time. Once we have a better understanding of the basic principles of message passing we would like to look at such phenomena as:

5.1 Robustness

Rather than having a predefined selection of rules to relax (Sondheimer & Weischedel, 1980), or using exhaustive backtracking, we believe that with message passing we can use the nature and context of errors to seek information from other actors on ways of circumventing the impasse.

5.2 Noun Groups

Oft-found noun groups such as "The staff of the Select Commission on Immigration and Refugee Policy" are fraught with perils for the unwary natural language processor. Examples like these led Gershman (1979) to conclude "Both linguistic and world knowledge are required for correct and efficient handling of noun groups." He went on to argue for "the advantages of the simultaneous application of both kinds of knowledge, without separating the process of understanding into syntactic and semantic stages." (p. 57)

5.3 Parallelism

The origins of the actor methodology are in an investigation of parallel processing. Kornfeld (1979) has pointed out an interesting phenomenon, "combinatorial implosion," in communicating systems. In his example,a parallel and communicating search algorithm dramatically reduces the time behavior, even when the algorithm is run in a pseudoparallel, time-sliced environment. One of our overwhelming interests is to find whether this behavior is manifest in language understanding systems written using the actor methodology.

6.0 BIBLIOGRAPHY

Gershman, A.V. Knowledge-Based Parsing. (Yale University Research Report #156.) New Haven: Yale University, 1979.

Hays, D.G. The field and scope of computational linguistics. Proceedings of the International Conference on Computational Linguistics. Debrecen, 1971.

Hendrix, G.G. Human engineering for applied natural language processing. Proceedings of the 5th International Joint Conference on Artificial Intelligence. Cambridge, 1977.

Hewitt, C. Viewing control structures as patterns of passing messages. (MIT AI

Memo 410.) Cambridge: MIT AI Laboratory, 1976.

Kornfeld, W.A. Using parallel processing for problem solving. AI Memo 561. Cambridge: MIT AI Laboratory, 1979.

Phillips, B. A model for knowledge and its application discourse analysis. American Journal of Computational Linguistics, 1978, Microfiche 82.

Schank, R.C. Conceptual Information Processing. New York: American Elsevier, 1975.

Sondheimer, N.K., & Weischedel, R.M. A rule based approach to ill-formed input. Proceedings of the International Conference on Computational Linguistics, Tokyo, 1980. Pp. 46-53. ;

Woods, W.A., & Kaplan, R.M. The Lunar Sciences natural language information system. (BBN Report No. 2265.) Cambridge: Bolt Beranek & Newman, 1971.

A PARSER WITH SOMETHING FOR EVERYONE*

Eugene Charniak
Dept. of Computer Science, Brown University

ABSTRACT

We present a syntactic parser, Paragram, which tries to accommodate three goals. First it will parse, in a natural way, ungrammatical sentences. Secondly, it aspires to "capture the relevant generalizations", as in transformational grammar, and thus its rules are in virtual one-to-one correspondence with typical transformational rules. Finally, it promises to be reasonably efficient, especially given certain limited parallel processing capabilities.

1. Introduction

Syntactic parsing in Artificial Intelligence (AI) has always had its share of controversies. Many in AI have seen in this work "much wasted effort"[4] and suggested that "the heavily hierarchical syntax analyses of yesteryear may not be necessary" [6]. At the same time, syntactic parsers have been attacked by those in linguistics as "devoid of any principles which could serve as even a basis for a serious scientific theory of human linguistic behavior" [2]. And, while psychologists have been kinder, any psychologist must be uncomfortable with theories which, if taken literally, would predict that people cannot understand ungrammatical sentences — a prediction which are false.

In this paper we will propose a parser, named "Paragram", which goes some way to answering this criticism. In particular:

1) The parser is "semi-grammatical" in the sense that it takes a standard "correct" grammar of English and applies it so long as it can, but will accept sentences which do not fit the grammar, while noting in which ways the sentences are deviant. Thus it will parse (1) while still using grammatical rules for subject/verb agreement to distinguish (2) from (3).

   (1) *The boys is dying.[1]
   (2) The fish is dying.
   (3) The fish are dying.

2) The rules of the parser are intended to capture the relevant generalizations about language in much the same way as a good transformational grammar. Paragram's rules are nearly in one-to-one correspondence with those proposed in some versions of transformational grammar.[2] Despite the fact that augmented transition network (ATN) parsers are based upon transformational grammar, when examined closely typical ATN grammars [7] seem to be far from the above ideal.

3) The parser is reasonably efficient, (0.3 seconds/word for a group of test sentences) and would be very efficient if implemented on a machine with limited parallelism, so that the rules of the grammar all test the input in parallel, but only one is actually applied (estimated .04 seconds/word). Efficiency aspects will not be discussed further in this paper.

2. Handling Ungrammatical Sentences

2.1. Why We Need a Deterministic Parser

Paragram is based upon Marcus's parser "Parsifal"[3]. We will explain Parsifal shortly, but first let us explain why we chose it as a starting point.

Probably the best known parser in AI today is Woods' ATN parser [7]. However it would not be possible to base a Paragram type parser upon the ATN parsing model. To see why this is so, we need only consider that when Paragram finds an ungrammatical situation, it must simply recognize it as such, and continue as best it can. ATN's simply do not work this way. When an ATN finds an ungrammatical situation it takes it as evidence that it made an incorrect decision earlier in the sentence, and hence backs up to find the correct path. So, consider

   (4) Jack sold the ball.
   (5) Jack sold Sue the ball.

Suppose that an ATN parser initially decides to parse "Sue" in (5) as a direct object, just like "the ball" in (4). When it gets to the second noun phrase in (5), "the ball", it has no way to handle it, and hence it backs up and tries making "Sue" into a dative which has been moved before the direct object. But suppose we had the ungrammatical sentence,

   *Jack sold Sue ball.

Here the ATN would back up as well, but to no avail, since there is no way to get a grammatical sentence out of this.[3]

In a deterministic parser (one which does not back up) the parser may assume that it has parsed everything correctly up to the point where it runs into trouble. Thus Parsifal knows where the trouble lies. It is this property which makes it an ideal starting point for Paragram.

2.2. Parsifal

Parsifal has two basic data structures, a stack and a buffer. The stack contains the sentence constituents on which it is still working. If a constituent is complete, then it must reside in one of two places: first, it may simply hang off some larger constituent. So at the end of a sentence there is only one item on the stack, the topmost **s** node, and everything else hangs off it. Second, Parsifal may have a complete constituent, but not know yet where it should go. Such constituents are put in the *buffer* which is a storage area of limited size. An obvious example would be an individual word (which is clearly complete). A less obvious example would be a noun phrase which, while complete, might be attached at any one of several places in the tree.

Rules in Parsifal are of the typical situation/action type. To decide if it is applicable, a rule will most often look to see what is in the buffer, although, with some limitations, rules may also look at the stack. Two positions in the stack are special, the bottom of the stack, which is named **c**, and the lowest sentence node in the stack, which is named **s**. To take a simple example, in Parsifal the rule for recognizing passive constructions is this:

```
(rule passive-aux        ;The rule is named passive-aux
  [= be] [= en] *        ;It looks at first two buffers
  Attach 1st to c        ;It puts the "be" on the bottom-
       as passive.)      ;most node of the stack.
```

---

*This is an extended abstract of a much longer paper by the same name, available from the author. My thanks to Graeme Hirst, who commented on the original paper. This research was supported in part by the Office of Naval Research under contract N00014-79-C-0592, and in part by the National Science Foundation under contract #MCS-8013689.

[1]Unless we explicitly indicate to the contrary, this and all other examples in this paper can be handled by Paragram. When an example is ungrammatical, Paragram will recognize it as such, but produce a reasonable "deep structure" anyway. Furthermore, it will indicate what in the sentence it did not like.

[2]We will be using a version of transformational grammar which was current in the late sixties. The primary reason for this choice is its familiarity. It should not be assumed that Paragram must necessarily use a grammar of this type.

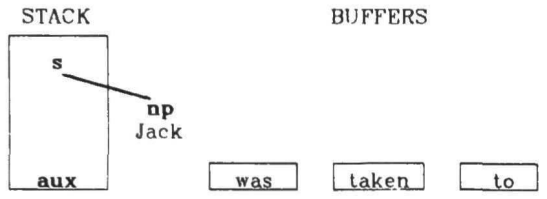
[3]Furthermore the time it takes an ATN to decide that the sentence is ungrammatical should go up roughly exponentially with the number of words. For some reason, those who tout ATNs as a model of human performance do not draw much attention to this "prediction".

The two square-bracket groupings indicate what the rule requires in the first and second buffer. In particular, the "=" indicates that what appears in the corresponding buffer must have the feature specified, such as being a form of the verb "be". Everything following the "→" is the action portion of the rule. These actions are specified in a language called "Pidgin", which is quite restricted but formulated to look like English.

Suppose we were applying this rule in the course of processing the sentence

Jack was taken to the house.

At the point where **passive-aux** is applicable, the state of the parser would be as follows:



Here the **np** "Jack" has been made a constituent of the top level sentence, but it is hanging off to the side to indicate that it is no longer on the stack, since it is a completed constituent. In the course of testing, Parsifal will see that "was" is a form of the verb "to be", while "taken" is an "en" form of the verb "to take" and thus the buffers match the rule test. At this point the action would be executed, which is to put the "was" on the **aux** which is currently the node Parsifal is working on. This will have the side effect of removing the "was" from the buffer, at which point the words further along in the sentence will move in to replace those which have been removed from the buffer.

However, not all rules of the grammar will actually be tested at any given point. Indeed most of the rules would be completely irrelevant; while parsing the auxiliaries of a verb we would hardly expect to find direct objects. To prevent Parsifal from even looking, each of its rules is found in one or more "packets" and only the rules which are in "active" packets will be tried. The active packets are those which are attached to the bottom node of the stack, **c**. Should this node be removed, the packets on the next higher node will be active. The idea is that if Parsifal is working on a noun phrase, then noun phrase rules will be active. Once Parsifal is done with it, it will be removed from the stack, and the rules on the next higher uncompleted constituent, say a verb phrase, will become active once more. Which packets are attached to a given node is explicitly controlled by the grammar rules themselves.

There are complications to this scheme, but this should due to give the reader a basic idea of how Parsifal works.

2.3. Parsing Ungrammatical Sentences

Naturally, Parsifal as currently constructed will only parse grammatical sentences. Should it be given an ungrammatical sentence, it will eventually come to a point where no rule applies, and it will simply give up.

Paragram differs from Parsifal in numerous ways, but allowing for ungrammatical input requires only a comparatively minor modification. Whereas Parsifal rules are tried sequentially until one works, active Paragram rules are to be thought of as being tested in parallel. Naturally, on current computers, they are really being tried sequentially, but it is useful to think of them as working in parallel. Furthermore, unlike Parsifal, the result of a test in Paragram is not a yes/no decision. Rather it is a numerical "goodness rating" which, the higher the number, the better the fit between the rule and the buffer/stack. Paragram then takes the rule with the highest number and runs it, allowing it to change the stack and buffers. It then repeats the process.

The goodness rating of a rule is the sum of the values returned by the rule's atomic tests. Each atomic test will add to the score if it succeeds, and subtract if not. No significance should be attached to the actual numbers. The basic idea is that the more tests succeeding, the higher the score, and failure is punished severely.

Now the crucial point in all of this is that for an ungrammatical sentence, the various ratings that we will get at the point of ungrammaticality will all be quite low, since none of the rules of grammar will exactly match the input. *Nevertheless, one rule must still have the highest score, and hence will apply, even though it does not really approve of the sentence as given.*[4] So, for example,

*The boys is dying.

will be given a low rating when Paragram starts to parse the auxiliary "is", at which point the best rule will be:

(rule subject-verb-agreement in parse-aux
[= verb] [test: The np of s agrees with 1st.] →
Create an aux. Activate build-aux.)

When applied to the above ungrammatical sentence this rule will have a poor goodness-of-fit rating since there is a match with **verb**, but the subject/verb agreement fails. Nevertheless, this is the best value at that point, so the rule is used anyway, and Paragram starts parsing the auxiliary verb, as intuitively it should. Note however that with sentences like

The fish is dying.
The fish are dying.

the above rule will succeed in each case, and in the process specify that the word "fish" is to be understood as singular and plural respectively. Some other ungrammatical sentences handled by Paragram are:

*Bill sold Sue book.
*Jack wants go to the store.

There are however, many ungrammatical, yet understandable, constructs which Paragram cannot currently handle. For example, extra constituents give it a problem. So more work needs to be done.

3. Parsing the Relevant Generalizations

The second goal set out for Paragram is that it capture appropriate generalizations about language in much the same way as a good transformational grammar. This has proved an elusive goal in parsing programs. The most well known of AI parsers, Woods's ATN parser, has been based upon transformational grammar, and Bresnan [1] points out that one could use the ATN framework to provide the link needed between her "realistic" grammar of English and an actual performance model of parsing. Nevertheless, while ATNs are inspired by transformational grammar, the single extended ATN grammar I have seen often required several special-case rules to handle what is a single rule in transformational grammar. While we will not pursue this point in any detail, let us take a single example, taken from the ATN grammar for English given in [7].

The rule of **there-insertion** in transformational grammar relates sentences like these:

There were barnacles on the ship.
Were there barnacles on the ship?
The ship on which there were barnacles sank.
The ship there were barnacles on sank.[5]

The statement of the rule is something like this:

[4] There is still the possibility of ties. However in practice this has not come up, and it can be argued that baring problems with the basic idea of deterministic parsing, ties should simply not occur.

[5] Stylistically this ain't so hot, but presumably it is grammatical. At any rate, the ATN grammar has a rule to handle it.

np(-def) exist-verb → There **exist-verb np** . .

This rule handles all cases of unstressed "there" (as opposed to the "there" in "There is Jack"). However, Woods's ATN has four separate rules for **there-insertion**, one for handling each of the above cases. So, despite the inspiration of transformational grammar, this ATN grammar has not done as well as one would like in the elegance of the rules it embodies [6]

Now in many respects, Parsifal does better. In particular, it needs only one rule of **there-insertion**. But this is not to say that all of Parsifal's rules are this elegant. *Indeed, of the 57 or so rules which I have looked at in depth from the detailed grammar at the end of Marcus' book [3], only twenty or so correspond to transformational rules.* Of the rest, they may be categorized into three groups, depending on the particular deficiencies which required them to appear in the grammar.

Miscellaneous problems. Six of the rules are needed because of various peculiarities of the grammar and the parser. While some of these are interesting in their own right, (particularly two, which deal with the rule of **raising**, a controversial rule in Linguistics) we shall say no more about them here.

Phrase Structure Rules. The majority of the 31 other rules, 21 in all, are there because Parsifal must have explicit rules in its grammar for the placement of phrase structure constituents. Thus, a rule like

s → np vp

is implemented by four separate rules in the grammar: one for creating an **s** when needed, one for attaching the **np** at the right spot, another for the **vp**, and finally one which says to stop parsing the **s**. Not only would it would be preferable to have a single rule, but the packet mechanism in Parsifal is really a phrase structure mechanism in disguise, thus these twenty one rules are redundant, at least in principle.

Paragram solves this problem by explicitly using phrase structure rules to handle packet switching, as well as replace many of the aforementioned rules. However, it has not proved possible as of yet to replace the rules which create new constituents of the appropriate type. This is because such rules typically differ widely from one another in what they are looking for in the buffer to clue them in that the new constituent is needed. These creation rules are currently the only rules which are not in one-to-one correspondence with typical transformational rules.

Wh-movement. Next, the ten remaining rules are involved in the implementation of the **wh-movement** rule, as used in

Who did Jack give the ball to?

Some of these rules are needed because Parsifal has two sets of verb-phrase rules. There is a normal set, which handles most verb phrases, but as soon as we must worry about gaps, we have a completely different set. This second set differ from the first in two ways. First, because they must worry about gaps, this second set continually checks the semantics to insure that it has not gone astray. In fact, there are cases where it is only semantics that can tell the parser what to do. For example

What did Bob give the girl?
Who did Bob give the book?[7]

Second, this second set contain many rules, each looking for a different configuration of things in the buffer which, in turn, will suggest where to locate the "gap" left behind by **wh-movement**.

Paragram does without all of this by making two changes in the parsing mechanism. The first is a technical change in the way the parser decides to postulate that a **wh** might have been moved from a particular location. The second, and perhaps more interesting change, is to avoid needing two sets of verb-phrase rules (with and without calls to semantics). Paragram *always* checks to see if a constituent is semantically reasonable before it will add it to the syntactic tree. Note that the user need not explicitly specify that a call to semantics is required here. Rather, Paragram automatically adds such calls to the testing section of any rule which adds a constituent to the tree [7]

4. Conclusion

While we have given no reason to take Paragram seriously as a model of human cognition, if we were to do so we would want answers to two important questions. First and foremost, how does syntax fits into the overall parsing process? Paragram essentially takes the conservative view that syntax is the initial mechanism which takes the word string and produces as its output some "deeper" representation that has properties that make it useful for the further pragmatic processing. It differs slightly from this however in that it requires that semantics be performed on a constituent before it can be attached to the tree.

Once we have decided to adopt a model in which syntactic analysis is done, and done as a separate process, we must then answer the second major question: what is the relation of the syntactic parsing process to standard "competence" models of syntax? Again, Paragram takes the old fashion view that this relationship is reasonably direct. Paragram argues that it may be possible to have a one-to-one correspondence between parsing rules and rules of grammar in a parser which is computationally efficient.

These views were, of course, very common in the sixties. They are less common now, in part because of various psychological results such as those of Slobin [5]. It is not my intent to try to refute the interpretation placed upon these results. Rather I hope that the existence of Parsers like Paragram can reopen the debate on these crucial subjects.

References

1. Joan Bresnan, "A Realistic Transformational Grammar," in *Linguistic Theory and Psychological Reality*, ed. M. Halle, J. Bresnan, and G. Miller, M.I.T. Press, Cambridge Mass. (1978).
2. B. Elan Dresher and Norbert Hornstein, "On Some Supposed Contributions of Artificial Intelligence to the Scientific Study of Language," *Cognition*, (4) pp. 321-398 (1976).
3. Mitchell P. Marcus, *A Theory of Syntactic Recognition for Natural Language*, M.I.T. Press, Cambridge, Mass. (1980).
4. Christopher K. Riesbeck, "Computational Understanding," pp. 11-15 in *Proceedings of the First Workshop on Theoretical Issues in Natural Language Processing*, ed. R. Schank and B. L. Nash-Webber, (1975).
5. Dan I. Slobin, "Grammatical Transformations and Sentence Comprehension in Childhood and Adulthood," *Journal of Verbal Learning and Verbal Behavior* 5 pp. 219-227 (1966).
6. Yorick Wilks, "An Intelligent Analyzer and Understander of English," *Communications of the ACM* 18(5) pp. 264-274 (1975).
7. William A. Woods, "An Experimental Parsing System for Transition Network Grammars," pp. 113-154 in *Natural Language Processing*, ed. R. Rustin, Algorithmics Press, New York (1972).

[6] It is important, however, to keep in mind the distinction between limitations in the parser and limitations in the particular grammar. It is possible that a more clever grammar writer might have avoided this problem. Indeed Woods has claimed (personal communication) that Parsifal is simply one kind of ATN. This may be so, but only in the uninteresting sense that both are one kind of Turing machine. In particular, the only ways I can see of simulating Parsifal with an ATN would completely ignore all of the built-in features that make ATNs what they are. At any rate, if it is true that Parsifal is one kind of ATN, then it must surely be the case that any improvement in Parsifal's grammar over a particular ATN grammar must only indicate deficiencies in the ATN grammar.

[7] There are two exceptions, due to problems in our handling of **dative-movement**.

Thought Sequences and

The Language of Consciousness

Benny Shanon
Department of Psychology
The Hebrew University, Jerusalem, Israel.

Consider the following episode, to be referred to as (1):

It is early evening. I am on my way back from the supermarket, and I pass through a small public garden. "It is a shame that I cannot pick some flowers for H., because it is dark," I think to myself. It occurs to me, however, that "but it is possible to buy" (flowers, that is), and I note that "she will enjoy me bringing her some". Consequently, I decide to "buy", I turn back and I go back to the supermarket in order to purchase some flowers.

The episode is the story of one thought sequence. Sequences of this type, I presume, are familiar to everybody. At various occasions we find that a thought is "passing in our heads": There is a trigger, several phrase-like expressions follow, and then one feels a break, an end. In the previous paragraph, the thought sequence is triggered by the sight of the bushes; it consists of four thought states (marked by inverted commas), and it terminates with a call for action. This form is general, thought sequences are cognitive entities characterized by beginning and end points which are well-demarcated and between which there is an ordered series of discreet steps that usually consist of linguistic-like thought expressions. The present paper is a brief report on a few patterns revealed in a comprehensive investigation of a large corpus of thought sequences I have collected over a period of several years. Such an investigation is interesting, I believe, because thought sequences constitute a natural cognitive kind, a genuine expression of the workings of the mind, not a reaction imposed on it by the demands of an artificial task (other such natural kinds are linguistic expressions, common-sensical inferences and dreams).

The corpus was collected by means of introspection. While in some paradigms of modern psychology, the method of introspection has been regarded as "the most fundamental of all the postulates of psychology" "what we have to rely on first and foremost and always" (James, in the Principles of Psychology), the method is commonly regarded as non-scientific and unreliable. Such a critical judgement, I think, stems from misappreciation of the appropriate usages of the method, and a failure to distinguish them from its abuses. The most notorious of all critiques of introspection have been raised in response to the research conducted in Würzburg at the beginning of the century. The members of the Würzburg school used introspection as a method for the direct solution of psychological issues. They believed that in order to find out how concepts are represented and processed, it is sufficient to look inside one's head and to observe. If one is perceptive, careful and experienced enough, one is bound to find a factual answer to the issue at hand (the Würzburgian subjects, recall, were the leading psychologists of their time). Introspection, however cannot be used for the direct solution of psychological issues and the generation of theories, but only for the collection of data. The analysis and modelling that ensue do not depend on introspection; in the present case they are conducted in a manner employed in the linguistic study of texts. As a consequence, the investigation is objective in the sense that it need not be confined to the study of one person, the one who had furnished the sequences. In fact, the study is based on the analysis of sequences furnished by several informants, and no significant individual variations were noted.

The use of introspection does require precautions. In order to safeguard against contaminations of memory and interpretation, all the sequences in the corpus were collected at the very time of their occurance. More important, however, are the analytical (not the technical) precautions. The employment of introspection sets limitations on the perspective by which the corpus may be examined. Clearly, there is no way to compare the thought expressions to any pure thoughts which underly them. Hence the analysis cannot deal with the structure of individual thought expressions, but only with the patterns which are exhibited by entire sequences These patterns are to be evaluated in terms of intrinsic coherence, not extrinsic correspondence. In fact, the present corpus exhibits both a systematic coherence and a completeness in that at present the probability for any new sequence to reveal a pattern not already specified in the model is close to nil.

That the topic of the investigation is sequences of thought expressions, not thoughts in any pure form, need not be viewed as a shortcoming. On the contrary the formulation of expressions in an inner code, which is very much like a natural language is itself a cognitive phenomenon. Clearly, not all thought processes are amenable to introspection: reaction time psychology, as well as psychoanalysis, attest to this fact. On the other hand, however, it is conceivable that the human mind would have been cognizant only of thought states which have practical ramifications, such as the call for action which terminates sequence (1). That the degree of resolution exhibited by thought sequences is of intermediate order is itself of cognitive significance Following this line of reasoning, the study of thought sequences by means of introspection cum linguistic analysis may be viewed not as the limited study of the shadows of thoughts, but as the genuine study of the language whose totality constitutes human consciousness.

From the present perspective the grammar of the language of consciousness is the set of mappings defined by the sequences of thought expressions. Local mappings define the relationship between two successive states in a sequence, but it appears that the history of the sequence is itself a determinant of its progression and that global mappings also need to be postulated in the grammar. For lack of space, only general patterns regarding the local mappings will be noted.

The different local mappings, it was found, constitute different patterns defined on a small number of parameters. Most significant of these is the level serving as the basis for the mapping. Local operations may be based on the relating of thought expressions by means of stored links, content, structure or particular symbols. The first of these bases defines associations. Associations relate items in stored representation by means of a link which is also stored. Associations vary according to the types of the representation they relate, the scopes associated with them. Representations may be lexical, semantic, episodic, phonological modality-specific, or motor; whereas scopes may involve one constituent, several constituents, or entire thought expressions. Traditionally, the term "association" has been used to refer to lexical-lexical mappings of a unitary scope. In the present corpus, however, instances of all possible associations were encountered. (2) for instance, is an example of an association relating one constituent in the perceptual domain to one entire episodic phrase.

(2) 0. Hearing a tune.
1. N. introduced me to this singer.

Content operations relate thought states on the basis of one's knowledge of the world. The items related by the mapping may themselves be either stored or generated. The operations may either supply further information regarding a given ite specify similar information regarding other items.

Formal operations relate thought expressions by form, not content, hence they are most useful in the generation of new information. They include: negation, interrogation, generalization, specification and conversion. While the terms noted are familiar from logic, the operations they denote are more general than the traditional formal ones. Generalization, for instance, is an entire family of operations, which vary according to the parameter and scope of both domain and range. (3) is a fragment in which two applications of generalization are noted; the domains and ranges of each appreciation are underlined:

(3) 1. to ask R. about it
2. to ask adults about it
3. to conduct experiments on adults.

The move from "ask about" to "conduct experiments", note, would not be classified as generalization in traditional treatises.

Lastly, operations whose basis is the particular symbol involve a shift in the perspective (reading or level) by which the given thought expression is processed. Here are two examples:

(4) 0. Eating spaghetti with a spoon.
1. Why isn't there a tool that will keep the noodles and let the liquid pass.
2. fork

In (4) the shift is from the definition of the symbol to its specification, whereas (5) the shift is from one reading of a given string to another.

(5) 0. (reading) I draw my finger across her forehead.
1. An image of myself drawing my finger across L.'s forehead.

The typology of local operations is only one face of the study of thought sequences. The richness of the thought machine, it appears, is a product not of the possession of a rich repertory of many, complex operations, but rather of the rapidity and flexibility of their dynamics. Most notably the following cluster of patterns will be noted:

a) Each constituent of a thought expression is a fuzzy designator, which may trigger a thought operation through its different aspects, some of which may be only partially specified.
b) The sequencing of mappings is driven by shifts of scope whereby a thought expression may be generated via one constituent, but serve as the basis for another expression via another.

c) Similarly, shifts of levels are encountered. Some shifts are recursive, whereby the very application of an operation serves as the trigger for the next state in the sequence.

d) Throughout, the application of operations involves interactions, both amongst themselves and with other psychological modules.

Together, these patterns suggest the metaphor exemplified by the following episode (for me, an habitual pattern):

I am at the top of a steep rocky slope. I have to descend it, but I am afraid. My solution is to run down as fast as I can. I jump from one stone to another, and by avoiding being stationed at any one of them, I manage to transverse the entire route through them.

Returning to the cognitive domain, this cascade pattern seems to be the one that enables not only the processing of given information, but also its generation de novo.

What conditions should a theory of consciousness meet?

Bernard J. Baars
SUNY Stony Brook

This paper is written in the conviction that theories of consciousness today are in the same situation that semantic theories were in a decade ago. At that time there was a widespread belief that semantics was an essentially insoluble puzzle, that much more data would have to be collected to constrain adequate theory. In the event, it turned out that the real obstacles were conceptual rather than empirical. Once the actual conditions to be met by an adequate semantics are specified, the theoretical options are vastly reduced. Further, conditions on semantic theory were not difficult to find: all such theories must be able to handle discourse reference, paraphrase generation, question answering, the detection of anomaly and contradiction, and the ability to resolve ambiguities at all levels of analysis. Until people looked at these actual constraints, no progress could be made, and the problem was treated as insoluble. Similarly, until we actually look at widely-accepted constraints on theory of consciousness, the topic will be treated as fuzzy, mystical, and insoluble.

What are the conditions for an adequate theory of consciousness? First, we can specify some pre-theoretical criteria:

1. An adaptive construct. Consciousness should be treated as a cognitive construct much like any other, with an adaptive information-processing function.

2. Relationship to other constructs. The proposed construct should be distinct from others, but explicitly related to perception, memory, intentionality, executive functions, automaticity, availability, the "internal monologue", the subjective observer, and especially attention.

3. What is UNconscious? A theory of consciousness should give a (principled) explanation of the difference between conscious and unconscious processes.

4. Respect for common-sense psychology. There is a world of difference between bootstrapping one's way beyond common sense and blindly ignoring it. Common sense is our starting point, and we cannot even ask about the nature of consciousness without it.

5. Empirical reference. Finally, a clear, empirically-based domain should be specified. Tables 1 and 2 show a number of contrasting pairs of claims about similar conscious and unconscious processes. These claims command a wide consensus among psychologists. The job of theory is then to fit some explanatory model to the constraints in the simplest possible way.

Table 1: Capability Constraints on a theory of conscious contents.

Conscious Processes	vs. Unconscious processors
1. Computationally inefficient.	Highly efficient in specialized tasks.
2. Great range, & relational capacity.	Limited domains & relative autonomy.
3. Apparent unity, seriality, & limited capacity.	Very diverse, parallel, and together have great capacity.

Table 1 shows some Capability Constraints, which indicate the capacities and limitations of conscious vs. unconscious phenomena. For example, if we use the term "computational efficiency" to mean the ability to work out some algorithm quickly and without error, it is clear that wholly conscious processes are not computationally efficient. Even simple addition or subtraction is performed slowly and with a good chance of error. It appears that the great bulk of fast and efficient processing is done by a large set of specialized processors. There is much neurophysiological evidence to this effect as well (Geschwind, 1979).

The great range and diversity of conscious contents seems to compensate for these efficiency limits. People can be conscious of virtually any energy pattern impinging upon any sensory system, down to single protons hitting single visual receptors. By means of conscious biofeedback, one can gain voluntary control over the actions of two-neuron spinal motor units. Etc. Or one can be conscious of events that require enormous cooperative activity between many millions of neurons.

Further, conscious contents always appear internally consistent at any one time, even if this consistency is spurious. This is in agreement with the fact that conscious capacity is limited and that the contents of consciousness appear serially. These facts belong together: If there must be internal consistency at any one time, then there is a clear capacity limit for incorporating mutually inconsistent contents, and such mutually exclusive contents must also appear serially.

In contrast to conscious processes, there is reason to think that unconscious processors can operate autonomously in their specialized domains without difficulty, because they are isolated from each other. They seem to be highly diverse, operating fast, efficiently, and in parallel. Taken as a whole, the set of all unconscious special-purpose processors has a very great capacity indeed.

These Capability Constraints have led me to associate consciousness with a well-known information-processing configuration: A global data-base, operating in a very large, distributed system. The global data base is essentially a central information exchange which permits otherwise autonomous specialized processors to interact with each other. Representations in the global data base are globally distributed, so that any one of a myriad specialists can respond to the global information, and some set of specialists can cooperate in return to create another global representation.

In this system, global processes are inefficient, slow, and error-prone because they require cooperation between different sets of specialists. Yet global information will have great range, diversity and context-sensitivity precisely because it involves interaction between many specialized processors in the system. Global information will show apparent unity, because inconsistent global representations will lead to competition between mutually exclusive specialists, which will cause the global representation to become rapidly unstable. There will thus be a narrowly limited capacity to display mutually competitive contents at any one time, and these will have to be displayed serially. In this way, all the Capability Constraints of Table 1 can be shown to apply to global representations in a distributed system. But this is not the whole story.

Table 2: Boundary constraints on the contents of consciousness.

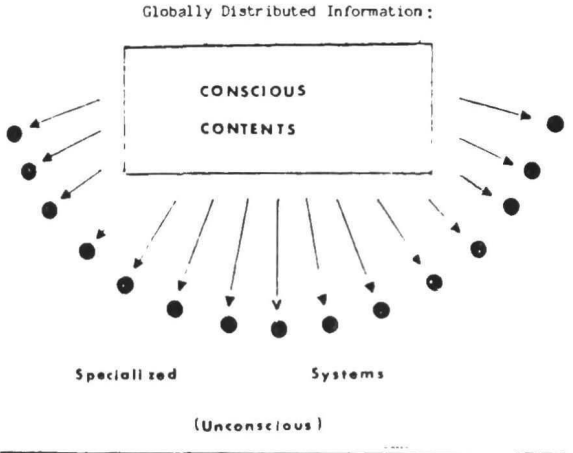
Conscious phenomena	Unconscious phenomena
Simultaneous cases:	
1. Percepts.	Context needed to organize percepts.
2. Input consistent with context.	Input inconsistent with context.
Diachronic cases:	
3. Percepts.	Pre-perceptual processes.
4. Any change in a habituated stimulus.	Habituated percept.

Consider the conscious-unconscious contrasts in Table 2, entitled Boundary Constraints --- a well-known set of facts about perception. Percepts are conscious, but the rapid pre-perceptual hypothesis-testing that is needed to establish the conscious percept is unconscious. Furthermore, the physical stimuli that lead to the conscious percepts are only conscious if they are defined within a stable set of contextual constraints. These contextual constraints are needed to provide the presuppositional background for the percepts, and they are not conscious. The physical stimuli themselves are not conscious either if the appropriate contextual background is missing. Further, percepts that are thoroughly analyzed drop out of consciousness (habituation and automaticity).

These facts imply that conscious contents involve more than just global information. They need to be stable and coherent , to accomodate the fact that rapid pre-perceptual processing is not conscious. Further, for a global representation to be conscious, it must be able to trigger widespread adaption in the nervous system --- once this adaptation has occurred, the conscious percept becomes unconscious, presumably because it has now become a part of the stable, presuppositional background. To put it all together, then: conscious contents must be global, coherent, and informative .

A first approximation to a system that fits these constraints is shown in Figure 1.

Figure 1

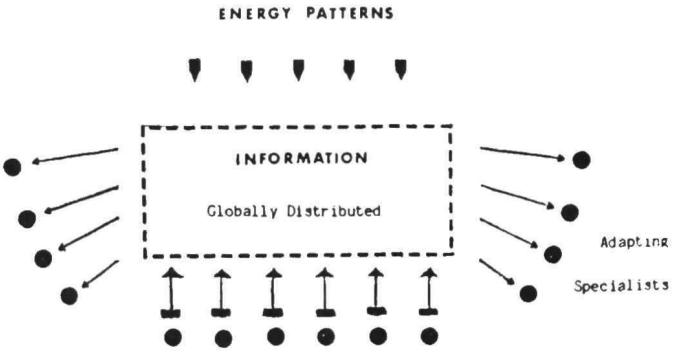


Note that conscious contents are globally available, but most detailed information processing is performed locally by a large set of specialized, distributed processors. The specialized processors maintain the processing initiative.

Conscious contents are not only global, but als coherent, because internally inconsistent contents would imply competition between different sets of specialists. Any active competition would rapidly remove the incoherent representation from the global data base.

Figure 2 presents a more refined system to fit the constraints set out in Tables 1 and 2.

Figure 2.



The global representation is now shown to be bounded by a set of stable contextual specialists which provide the presuppositions within which the global representation is defined. However the contextual specialists are not themselves conscious. Other specialists are in the process of adapting to the global representation. Once adaptation is complete, the global representation fades from consciousness; it becomes part of the contextual background, though it continues to constrain other conscious contents.

This final model can handle all the major empirical constraints on theory listed above: Conscious contents must be global, internally coherent, and informative (i.e., able to trigger widespread adaptation).

This theoretical approach is treated in much greater detail in Baars (in press) and Baars and Mattson (in press).

Baars, B.J. Conscious contents provide the nervous system with coherent, global information. In R.J. Davidson, G. Schwartz, and D. Shapiro (eds.) Consciousness and self-regulation Vol. III , NY: Plenum, in press.

Baars, B.J. and Mattson, M.E. Consciousness and intention: A framework and some evidence. Cognition & Brain Theory , in press.

Geschwind, N. Specializations of the human brain. Scientific American, 241 (3) .1979,180-201.

THE ROLE OF THE OBSERVER IN COGNITIVE SCIENCE.

Richard Hammersley.*

The current philosophical basis of cognitive science leads to problems in the theories of cognition proposed. An alternative basis, using phenomonology, may be more viable.

Western metaphysics has been divided between dualists and monists. Both equate self and mind. Cognition is the relation between the self or mind and the world or the rest of the world.

Cognition requires a world and a mental representation of that world. Unfortunatly we cannot be sure of the world. Any variety of realism requires some representation of the world. Theorising becomes the relation of on model to another. Because of our prejudices we tend to ascribe one of these models reality. The subjects of our theorising are taken to possess representations of this world. We deduce the nature of their representations, and the processes necessary to gain these representations upon the assumption that our "observers world" is real. In fact what we take to be properties of the subject may be imposed by the nature of the assumptions within the observer's world.

Phenomonology equates the world with the self (e.g. Merleau-Ponty 1962, 1964). My world is myself. Mind is a part of myself which I have learned to contrast to the real. Rather than trying to control one's own subjectivity, as an observer, one must understand it. There is no real world to base one's cognitive models on. Instead one must consider the relationship between worlds. What cannot be communicated cannot be studied. As researchers we devise many ways for subjects to communicate with us. It is a mistake to ignore their role in this communication.

As cognitive scientists we should consider the worlds of subjects. To do this we must abandon our privileged position as "objective" observers. We have struggled to objectify our analyses of behaviour and other manifestations of cognition, this attempt is spurious. We will only progress by devising ways of recording and relating subjective worlds.

This may look as if it will be very difficult, and further, non-objective. The remainder of this paper argues that 1. Current cognitive science theorising is already non-objective and that this poses a serious problem. 2. This problem must be faced by making the role of the observer explicit in theory.

Let me provide examples of 1. I hope that they will be clear enough to allow you to generate others. Let me take a psychological example, a linguistic example and an A.I. example. For psychology, consider the concept of imagery in particular and mental represen tation in general. The intial interest in imagery was fairly simplistic, early studies sought to discover if there were any apparent performance correlates of a mental image. Reports of imagery were correlated with performance in one way or another. It has since become clear that reporting an image correlates with some analogue kinds performance especially in terms of "distance" traversed over time. It has also become clear that reported imagery is not necessary for such effects (Friedman 1979) and that use of "the image" is not like use of an object (Hinton 1980, Richman and Mitchell 1979). So we now have representations which are like images, but clearly cannot correlate precisely with imagery reports. The problems of interpreting the original observed performances have not been solved. It is not clear what the basis of imagery reports or time differences in processing are. The assumptions that time differences represent real processing and that reported imagery has some imaginal (e.g. analogue) basis are part of the observer's or theorist's world. I suggest that the observers have found evidence which supports their world. They have not found evidence which would convince their more cynical colleagues. By wading into such exciting problems as imagery, without being clear what they believe an image could be, cognitive psychologists have produced a lot of experiments which are uninterpretable. You can like them or dislike them but is is unclear what they have shown. The image as a concept may be a useful part of the observer's world, it is not clear, and cannot be without the distinction, whether it is a useful part of the subject's world, or even a functional entity within the processing system.

Space prevents me going into detail for the other examples. I'll just name them to annoy and tease those who believe in them. In AI the notion of a representation of knowledge seems to be a formalisation of the observer's knowledge, it is not a theory of knowledge in itself. In linguistics the notion of a language is part of the observer's/community world. It is not clear that it is an essential part of linguistic theory. Need one assume the object of study to understand it? It would be better to derive it.

How will considering the role of the observer help to solve such problems? For a start, it becomes clear that there are no "bald facts" which we all agree about as decent and upright scientists. For example, pro-imagery people seem to believe in the reality of images, linguists believe in language. For some purposes these beliefs may be useful, they cannot be validated by assuming them. Nor can the existence of such things be proven. If we want to understand the mind of a subject, rather than simply capturing behaviour in a descriptive way, then we have the problem that their beliefs and ours may be different. Supposing one found "imagery effects" in a person who reported no imagery. The current way of understanding this would be to claim that the person had an unconscious image. It is important to recognise that this is a claim about our world, not the subjects. For example it might be futile to tell such a subject to "picture" something, other ways of enabling "imagery" might be necessary. The ideal, if we wish to study the minds of others, would be some model which both observer and subject could agree on, or at least be made to agree on.

One role of science is to establish such an agreement within a scientific community. For example, Human Information Processing has certain agreements about how to consider psychology. A critical notion is that of information. Recently, we have seen criticisms by Gibson (e.g 1966) and Turvey (e.g.1977) of the assumption that there is some absolute way of defining units of information in the enviroment. Their alternative becomes subject relative. I suggest that this is too simple. The "primitive units of information" in the enviroment will depend upon the subject, the enviromental context, the historical context of the subject and, not least important, the way in which the observer characterises the problem of interest. That is, the notion of invariance only has meaning within a system which includes the observer, the subject and the enviroment. An affordance, like edible, or a feature, like straight line, can only be specified in this manner. The fact that they appear "real" invariants to us is due to our ignorance about our participation in this process. You may protest that by including our participation in science, we are liable to destroy the objective and repeatable nature of science. Clearly the functional utility of several sciences, or better, systems of scientific communication, is great. It is not clear that this is the case in cognitive science. I am proposing that cognition be regarded as a relation between the world of the observer and the world of the subject, rather than as a static system which can be charted and understood as processing within a single world. I believe that this will overcome the problems with using phenomonology as the basis merely for introspection or some uncritical description of people's beliefs. I am suggesting using two phenomonologies at once. The relation between those worlds will serve as the basis for cognitive science, rather than attempting to base one world in

the other, as is presently done. As a glib summary to remember: What is real is not important, what the subject and the observer believe to be real is critical.

References.

Friedman A. 1978. Memorial comparisons without the mind's eye. Journal of Verbal Learning and Verbal Behaviour. 17. 427-444.

Gibson J.J. 1966. The Senses Considered as Perceptual Systems.

Hinton G. 1979. Some demonstrations of the effects of structural descriptions in mental imagery. Cognitive Science. 3. 231-250.

Merleau-Ponty M. 1962. The Phenomonology of Perception. RKP.

Merleau-Ponty M. 1964. Philosophy and Sociology. In Signs. Northwestern University Press.

Richman L.L. and Mitchell D.B. 1979. Mental travel: some reservations. Journal of Experimental Psychology: Human Perception and Performance. 5. 13-18.

Turvey M.T. 1977. Preliminaries to a theory of action with reference to vision. In Perceiving Acting and Knowing. Ed. R. Shaw and J.B. Bransford. LEA.

*Program in Cognitive Science. UCSD C-009, La Jolla, Ca 92093.

Are We Ready for a Cognitive Engineering?

S. K. CARD, *Xerox Palo Alto Research Center* AND
A. NEWELL, *Carnegie-Mellon University*

It is an interesting irony that psychology is often criticized for its impracticality and its concern with minutiae while at the same time the crucial problems impeding many engineering developments are problems of psychological performance. Nowhere is this more true than in the area of human-computer interaction, whether it is a matter of automation in the cockpit of advanced aircraft or in the tasks of office workers. In both cases it is the lack of understanding of the determinants of psychological performance in the user that currently sets limits on the use of technology.

Yet, anyone who has had the task of trying to obtain from the literature psychological guidance for the design of an interactive computer system is aware of the great frustrations engendered by the jumble of empirical results and micro-theories, tightly bound to experimental paradigms, which he finds. It is not that psychology has no information to offer. Indeed, while the literature admittedly contains quantities of ill-founded trivia, there has come to be established in cognitive psychology a solid set of verified facts about the working of the human mind. But these facts are difficult to retrieve and to apply in new circumstances. This is true partially because in psychology, unlike some other fields, there has been insufficient effort devoted to codifying and condensing the knowledge learned into simple forms, usable by workers with other specialties. The codification and simplification of established facts is important to the progress of science if the results from one specialty are to become the tools of another (Latour and Woolgar, 1979), if the results of cognitive psychology are to coalesce into a science base for cognitive engineering.

But are we ready for a cognitive engineering? Is it now within our reach to create a systematic methodology for designing machines optimized for human cognitive performance? There is certainly the engineering need and there are also the promising developments in cognitive psychology, but it is a great mistake to think that, by merely listing a miscellaneous collection of results, cognitive psychology is thereby rendered usable to support a discipline of cognitive engineering. What is required is a more radical departure from what is usual in psychological research. Whereas the point of experimental manipulations is often to discriminate between competing theories no matter how small the discrimination, by contrast, in a psychology useful for engineering design, small differences are lost in the many approximations always necessary. What is important is the ability to do task analysis (determining the specific, rational means of accomplishing various goals), calculation (zero-parameter predictions of behavior capable of parametric variation), and approximation (simplification of the task and of psychological theory).

In the remainder of this paper, we wish to suggest a way in which numerous results from cognitive psychology might be included in a single model, usable by computer system designers and others. Though limited, this model (which we shall dub the Model Human Processor) does make it possible to calculate predictions of user performance, albeit of an approximate kind. The purpose of the model is not to provide a precise description of what is in the head, but to provide an economical and sufficient basis for applied analysis.

THE MODEL HUMAN PROCESSOR

The Model Human Processor can be described by (1) a set of [illegible] together with (2) a set of "principles of operation" (Figure 2). The principal properties of the processors and memories are summarized by a small set of parameters. A similar technique has proved successful for simplifying the analysis of electronic information-processing systems (see Siewiorek, Bell, and Newell, 1981). The memory parameters used in the model are are

- μ, the storage capacity in items,
- δ, the decay time (half-life) of an item, and
- κ, the main code type (iconic, acoustic, visual, semantic).

The only processor parameter used is

- τ, the cycle time.

The complete model is elaborated and argued in Card, Moran, and Newell (in preparation). Here, we wish only to give an illustration of how a single parameter τ from the model can be used to support system engineering analysis.

According to the Model Human Processor, the mind is comprised of three partially coupled processors, the Perceptual Processor, the Cognitive Processor, and the Motor Processor, each with a similar cycle time, derived from the literature.

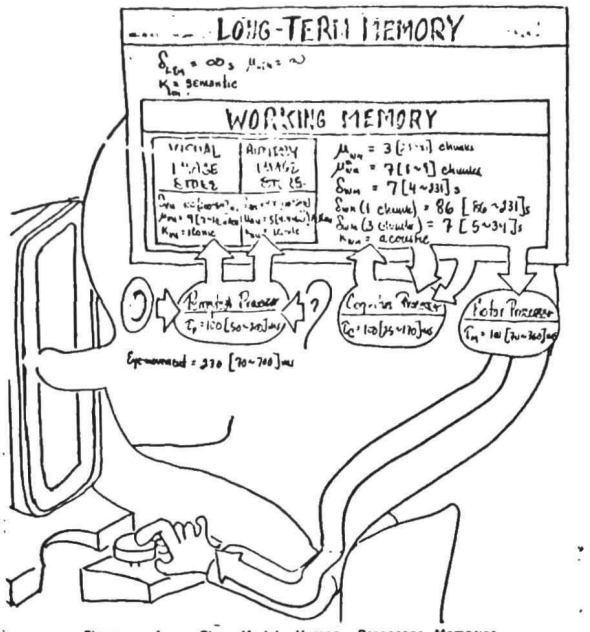


Figure 1. The Model Human Processor—Memories, processes, and Basic Principle of Operation.
Sensory information flows into Working Memory through the Perceptual Processor; Motor programs are sent out through activation of chunks in Working Memory. Working Memory consists of activated chunks in Long Term Memory. Visual Image Store and Auditory Image Store can be thought of as special activations of the experiential and analogic properties of visual and auditory memory. The basic Principle of Operation of the Model Human Processor is the Recognize Act cycle of the Cognitive Processor.
On each cycle the contents of Working Memory activate actions associatively linked to them in Long Term Memory, which in turn, modify the contents of Working Memory.

P1. Variable Perceptual Processor Rate Principle. *The Perceptual Processor cycle time τ_P varies inversely with stimulus intensity.*

P2. Encoding Specificity Principle. *Specific encoding operations performed on what is perceived determine what is stored, and what is stored determines what retrieval cues are effective in providing access to what is stored.*

P3. Discrimination Principle. *The difficulty of memory retrieval is determined by the candidates that exist in the memory, relative to the retrieval clues.*

P4. Variable Cognitive Processor Rate Principle. *The Cognitive Processor cycle time τ_C is shorter for greater task demands and increased information loads; it also diminishes with practice.*

P5. Fitts's Law. *The time T_{pos} to move the hand to a target of size S which lies a distance D away is given by $T_{pos} = I_M \log_2(2D/S)$, where $I_M = 100$ms/bit[70~120ms/bit].*

P6. Power Law of Practice. *The time T_n to perform a task on the nth trial follows a power law: $T_n = T_1 n^{-r}$, where $r = .4[.2 \sim .6]$.*

P7. The Uncertainty Principle. *Decision time T increases with uncertainty about the judgement or decision to be made: $T = I_C H$, where H is the information-theoretic entropy of the decision and $I_C = 150$ ms/bit [0~157 ms/bit]. For n equally probably alternatives (Hick's Law), $H = \log_2(n+1)$. For n alternatives with different probabilities of occurring p_i, $H = p_i \log_2(1/p_i + 1)$.*

P8. Rationality Principle. *A person acts so as to attain his goals through rational action, given the structure of the task and his inputs of information and bounded by limitations on his knowledge and processing ability:*
Goals + Task + Operators + Inputs + Knowledge + Process-limits → Behavior.

P9. The Problem Space Principle. *The rational activity in which people engage to solve a problem can be described in terms of (1) a set of states of knowledge, (2) operators for changing one state into another, (3) constraints on operator movement, and (4) control knowledge for deciding which operator to apply next.*

Figure 2. The Model Human Processor—Additional Principles of Operation.

τ_P = 100 ms [50~200]
τ_C = 100 ms [25~170 ms]
τ_M = 100 ms [70~360 ms].

For some tasks (pressing a key in response to a light) the system must behave as a serial processor. For other tasks (typing, reading, and simultaneous translation) integrated, parallel operation of the three subsystems is possible, in the manner of three pipelined processors: information flows continuously from input to output, with a lag showing that all three processors are working simultaneously.

Suppose a stimulus impinges upon the retina of the eye at time t 0. At the end of one Perceptual Processor cycle, $t = \tau_P$, the image is assumed to be available in the Visual Image Store and the human able to see it. Shortly thereafter, a recognized, symbolic, acoustically- (or visually-) coded representation of at least part of the Visual Image Store contents is assumed to be present in Working Memory. In truth, this description is an approximation since different information in the image becomes available at different times, much as a photograph develops, and a person can react before the image is fully developed or he can wait for a better image, depending on whether speed or accuracy is the more important. According to the model, perceptual events occurring within a single cycle are combined into a single percept if they are sufficiently similar.

Once in Working Memory, information is processed by the Cognitive Processor's *recognize-act cycle*, analogous to the fetch-execute cycle of standard computers. On each cycle, the contents of Working Memory initiate associatively-linked actions in Long-Term Memory ("recognize") which in turn modify the contents of Working Memory ("act"), setting the stage for the next cycle. Plans, procedures, and other forms of organized behavior can exist, but these are built up out of an organized set of recognize-act cycles.

Consequent to a decision to act by the Cognitive Processor, the action itself is controlled by the Motor Processor. Contrary to casual appearances, movement is not continuous, but consists of a series of discrete micromovements, each requiring one Motor Processor cycle of about τ_M = 100 ms. The feedback loop from action to perception is sufficiently long (300~500 ms) that rapid behavioral acts such as [illegible] must be composed of preprogrammed Motor Processor cycles.

CALCULATING HUMAN PERFORMANCE

We now illustrate how the portion of the Model Human Processor we have described can be used to calculate answers to problems of human performance. To address uncertainties in the parameters of the model, we define three model versions: one in which all the parameters listed are set to give the worst performance (*Slowman*), one in which they are set to give the best performance (*Fastman*), and one set for a typical performance (*Middleman*). Sensitivity analyses to see if the results of interest are affected by the true values of the parameters can be performed by making calculations for both Slowman and Fastman.

Example 1. Morse Code listening rate.

> What is the maximum rate, in words/min, at which Morse Code may be perceived? (Assume the old sort of American telegraph where dots and dashes are made by the clicks of the armature of an electromagnet, dots being distinguished from dashes by a shorter interval between armature clicks.)

Solution. If a dash takes less than one Perceptual Processor cycle τ_P = 100 ms, there would be no way to tell it from a dot since, according to the model, durations smaller than τ_P cannot be perceived. Similarly, if two dots occur closer than about 100 ms they would be perceived as the same dot. So a reasonable approximation of the fastest rate is 100 ms/dot, 200 ms/dash, and 100 ms between letters.

If the probabilities for the various letters in English are multiplied by the calculated time per letter according to the above rule, we calculate a mean time of 459 ms/letter. Assuming 4.8 char/word (the value for Bryan and Harter's, 1898, telegraphic speed test),

Max reception rate = (.459 s/letter × 4.8 letter/word) + .200 s/word-space
= 2.4 s/word = 25 words/min. ∎

This number is in the range quoted by Bryan and Harter (1898) for very good, experienced, railroad telegraphers, 20~25 words/min.

An upper bound on the maximum rate is given by a Fastman calculation:

Max rate = (100 ms/50 ms) × 25 words/min
= 50 words/min.

Of course this calculation does not guarantee that 50 words/min is actually possible, only that it would be surprising if anyone were to be faster. In fact, the fastest performance known to Bryan and Harter was close to this rate, 49 words/min.

Example 2. Reaching to a button

> Suppose a user needs to move his hand D cm to reach a button S cm wide on a calculator (He cannot reach it by touch). How long will the movement require?

The movement of the hand, as we have said, is not continuous, but consists of a series of micro-corrections each with a certain accuracy. To make a correction takes at minimum one cycle of the Perceptual Processor to observe the hand, one cycle of the Cognitive Processor to decide on the correction, and one cycle of the Motor Processor to perform the correction, or $\tau_P+\tau_C+\tau_M$. The time to move the hand to the target is then the time to perform n of these corrections or $n(\tau_P+\tau_C+\tau_M)$. Since $\tau_P+\tau_C+\tau_M \simeq$ 300 ms, n is the number of roughly 300 ms intervals it takes to point to the target.

Let X_i be the distance remaining to the target after the ith corrective move and X_0 (= D) be the starting point. Assume that the relative accuracy of movement is constant, that is, that $X_i/X_{i-1} = \alpha$, where α (< 1) is the constant error. On the first cycle the hand moves to

$$X_1 = \alpha X_0 = \alpha D.$$

On the second cycle, the hand moves to

$$X_2 = \alpha X_1 = \alpha(\alpha D) = \alpha^2 D.$$

On the nth cycle it moves to

$$X_n = \alpha^n D.$$

The hand stops moving when it is within the target area, that is when

$$\alpha^n D \leq \tfrac{1}{2}S.$$

Solving for n gives

$$n = -\log_2(2D/S) \;/\; \log_2 \alpha.$$

Hence the total movement time T_{pos} is given by

$$T_{pos} = n(\tau_P+\tau_C+\tau_M)$$
$$T_{pos} = I_M \log_2(2D/S) \qquad (1)$$
where $I_M = -(\tau_P+\tau_C+\tau_M) \;/\; \log_2 \alpha.$

Equation 1 is called Fitts's Law (this derivation based on Keele, 1968). It says that the time to move the hand to a target depends only on the relative precision required, that is, the ratio between the target's distance and its size.

The constant α has been found to be about .07 (Vince, 1948), so I_M can be evaluated:

I_M = −300 ms/$\log_2$(.07) bits
= 78 ms/bit.

Results from various experiments give values in the in the I_M = 70~120 ms/bit range.

Example 3. Reaction time.

> The user is presented with two symbols, one at a time. If the second symbol is identical to the first, he is to push the key labeled Yes, otherwise he is to push No. What is the time between signal and response for the Yes case?

Solution. The first symbol is presented on the screen where it is observed by the user and processed by his Perceptual Processor giving rise to associated representations in the user's Visual Image Store and Working Memory. The second symbol is now flashed on the screen and is similarly processed. Since we are interested in how

long it takes to respond to the second symbol, we now start the clock at 0. The Perceptual Processor processes the second symbol to get an iconic representation in Visual Image Store and then a visual representation in Working Memory, requiring one cycle, τ_P. If not too much time has passed since the first symbol was presented, its visual code is still in Working Memory and the Cognitive Processor can match the visual codes of the first and second symbols against each other to see if they are the same. This match requires one Cognitive Processor cycle, τ_C. If they match, the Cognitive Processor decides to push the Yes button, requiring another cycle τ_C for the decision. Finally, the Motor Processor processes the request to push the Yes button, requiring one Motor Processor cycle τ_M. The total elapsed reaction time, according to the Model Human Processor, is

$$\begin{aligned} \text{Reaction time} &= \tau_P + 2\tau_C + \tau_M \\ &= 100\ [50\sim200] + 2\times(100\ [25\sim170]) \\ &\quad + 100\ [70\sim360]\ \text{ms} \\ &= 400\ [170\sim900]\ \text{ms}. \end{aligned}$$

This analysis could be repeated for the case ("name match") where the user is to press Yes if the symbols were both the same letter, although one might be in upper case, the other in lower case. Here, an extra Cognitive Processor cycle is required to get the abstract code for the symbol (Computed reaction time = 500 [195~1070] ms). Likewise, if the user were to press Yes when the symbols were only of the same class ("class match"), say both letters, yet another Cognitive Processor cycle would be required (Computed reaction time = 600 [220~1240] ms).

Experiments have been performed to collect empirical data on the questions presented in these examples. The finding is that name matches are about 70 ms slower than physical matches and that class matches are about 70 ms slower yet, a number in line with our 100 ms [25~170 ms] value for τ_C.

The forgoing examples start from task analysis of a problem and proceed through approximation and calculation to make predictions of human performance of the sort that might be used in an engineering analysis of cognitive behavior. Although it is hoped that the Model Human Processor itself will be useful for engineering, the real point is in the spirit of the enterprise: that knowledge in cognitive psychology and collateral sciences is sufficiently advanced to allow the analysis and improvement of common mental tasks. In short, that a cognitive engineering is now feasible, provided there is a disciplined understanding of how the knowledge must be structured to be useful. Of course, suggestions for improvement to the present model will occur all around. But then that is part of the idea and the challenge: to use the Model Human Processor as a framework into which new research results and insights can be fit in a way amenable to use in the cognitive engineering of practical mental tasks.

REFERENCES

Bryan, W. L. and Harter, N. (1898). Studies in the physiology and psychology of the telegraphic language. *Psychological Review 4*, 27-53.

Card, S. K.; Moran, T. P.; and Newell, A. (in preparation). *The Psychology of Human-Computer Interaction*. Hillsdale, N. J.: Erlbaum Associates.

Keele, S. (1968). Movement control in skilled motor performance. *Psychological Bulletin*, 387-403.

Latour, B. and Woolgar, S. (1979). *Laboratory Life: the Social Construction of Scientific Facts*. London: Sage Publications.

Siewiorek, D.; Bell, G.; and Newell, A. (1981). *Computer Structures*. New York: McGraw Hill.

Vince, M. A. (1948). Corrective movements in a pursuit task. *Quarterly Journal of Experimental Psychology*, *1*, 85-103.

Knowledge Structures of Computer Programmers

Beth Adelson
Harvard University

Three research questions are addressed here: what are the knowledge structures of novice and expert programmers; what principles underlie these structures; and how is the knowledge used? The results of the two experiments described here indicate that novices form syntactically based representations of programs which are concerned with the details of how the program functions while experts form more abstract conceptually based representations which are concerned with what the program does. The results of these experiments suggest that the representation used in a task constrains performance and that new representations develop with expertise.

Experiment 1

In this experiment Novice and Expert computer programmers perform a multi-trial free recall task (MTFR). The stimulus set consists of lines of programming code which can be organized either syntactically or conceptually. The clusters found in the recall protocols of each group will suggest the nature of the knowledge structures of each group.

Method.

Subjects. Five Novice and five Expert programmers formed the two groups of subjects.

Stimuli. The organization of the stimulus set is central to this experiment. The stimuli consisted of 16 lines of code in Polymorphic Programming Language ("PPL"is a language similar to APL) which could be organized either syntactically or procedurally.Under the procedural classification the items (lines of code) can be organized into three programs: the first is a sorting routine, the second and the third are random sampling routines. The second basis for the classification of the stimulus set is syntactic, with syntax being used here the way it is used to describe natural language. Different control phrases of the computer language act as different parts of speech in that they expect certain other kinds of control phrases to precede or follow them.There are five different syntactic categories present in the stimulus set.

Procedure. Each subject saw all of the items, one at a time and then had to recall as many of the items as possible. This procedure was repeated nine times.The stimulus set was presented in a different random order on each trial. The presentation of items was not blocked either by program or by syntactic category. Neither group knew what the organization of the stimulus set was although both groups were familiar with the syntax of the language as well as the concepts behind the three programs.

Results and Discussion.

The Novices recalled significantly more than the Experts, they also had larger chunks and more consistent subjective organization. In order to look at the organization underlying these quantitative results multi-dimensional scalings and hierarchical clusterings were done on the recall protocols of each group. In Figure 1 we see the two-dimensional scaling solution

--- --- --- ---

Insert Figure 1 Here.

--- --- --- ---

of inter-item similarity for the Novice group. The points in the solution are the

items, they are labelled by syntactic category. The procedural classification of each item is given in parentheses. For the Novices the items cluster by syntactic category; the Assignment statements, (the A's), cluster in the first quadrant with the Conditional IF statements (the I's) around them. The iteration or FOR statements (the F's) cluster in the second quadrant, the function Headers (the H's) cluster in the third quadrant and the RETURN statements (the R's) cluster in the fourth quadrant.

--- --- --- ---

Insert Figure 2 Here.

--- --- --- ---

The two-dimensional scaling solution for the Expert group is presented in Figure two. Here we see the same set of items, but this time they cluster by program. In the labelling of the items the first number for an item represents the program that it belongs to, the second number for an item represents its positon in the program. For example item 2.0 is the first item in program two. The items from program one, those labelled 1.0 through 1.4, are in the first quadrant. The items from program 2, those labelled 2.0 through 2.4, are in the third quadrant and the items from program three, those labelled 3.0 through 3.5, are in the fourth quadrant. The syntactic classification is given here in parentheses, it does not capture the clusters of the Experts the way it did the clusters of the Novices. The original recall protocols also showed that each of the Expert subjects recalled all of the lines from all three programs in the order in which they would have been evaluated in a running program,this suggests that the Experts are using their knowledge of the serial nature of computer programs to organize the items within a cluster. It appears that the clusters of the Expert subjects are more abstract and conceptual than the syntactic clusters of the Novice subjects. Here as in other skilled problem solving domains we find that the chunks of the experts are based on the functional principles of the skill; items are categorized as members of one procedure or another and are then organized serially within those procedural categories. The chunks of the Novices however, are not functionally based, they are syntactic in nature. In addition, the recall protocols of the Novices showed no internal organization of the items other than the grouping by syntactic category, that is within the Novice categories there was no regularity as there was in the Expert categories.

Experiment II

The results of the first experiment suggested that the organization of the Novices was syntactic while the organization of the Experts was functional. It is possible then, that when given whole programs to comprehend experts form an abstract representation that is concerned not with the specific mechanics of a given algorithm but only with *what* it is that the program does. It is also possible that a representation formed by a novice, because it is synactically based would be more concerned with *how* the specific program functioned. The second experiment presented here was designed to check out the possibility that during program comprehension novices form representations of how programs function, while experts form representations of what programs do.

The strategy used to look at this possibility is to have both groups form and use

both types of representations and then to look at the resulting performance in each situation.

Method

Subjects. The subjects were a group of Novice and a group of Expert programmers.

Stimuli. The stimuli consisted of eight PPL programs with two types of flow charts and two types of questions for each of the eight programs. The flow charts consisted of Low Detail flow charts which described what the program did and High Detail flow charts which described how the program functioned. The questions consisted of Low Detail questions which asked a question about what the program did and High Detail questions which asked a question about how the program functioned.

Procedure. In order to encourage subjects to form a representation at one level of abstraction or another subjects were first shown a flow chart of a program at a given level of abstraction. They were then shown the actual program along with a question about the program. After seeing the program itself, along with the question the subjects answered the question . Level of detail of the flow chart was combined with level of detail of the question to form four conditions for each group of subjects; two "congruent" conditions and two "non-congruent" conditions. In the two congruent conditions subjects saw either a Low Detail flow chart and then a Low Detail question or a High Detail flow chart and then a High Detail question. In the two "non-congruent" conditions subjects saw either a Low Detail flow chart and then a High Detail question or a High Detail flow chart and then a Low Detail question.

Two times were recorded: Comprehension time, that is the time it took the subject to say that s/he understood the flow chart well enough to go on and study the program and the question together and Question time, that is the time interval between seeing the program with the question and being able to write down the answer.

A "Code Only" control conditon was included in which subjects saw the program itself immediately and then saw the same program again with the question.

Results and Discussion

I. Comprehension Time

(Interval between seeing the flow chart and understanding it)

Comparing the results of the Low Detail, High Detail and Code Only conditions, subjects understood the Low Detail flow charts more quickly than the High Detail flow charts and the High Detail flow charts more quickly than the code alone. The flow charts do seem to aid comprehension and it appears that they do so by organizing the information for the subjects rather than by just reducing the information since the High Detail flow charts, which had more units on the average than the code were understood more quickly than the code alone in the Code Only control condition (Units are boxes in the flow chart or single lines in the code.)

II. Question Time (Interval between seeing the question and being able to answer it)

A. Non-congruent Conditons (here the level of detail of the flow chart and the question do not match)

The results of the non-congruent conditions give us information about each group's most natural level of abstraction. The

rationale behind this is that the flow charts prepare the subjects to represent the code at one level of abstraction or another. If the level then turns out to be inappropriate for answering the question subjects will be substantially slowed down if the appropriate level is also not the level which they usually use. On the other hand if the appropriate level is the level which they usually use they should still be able to quickly form a representation at that level.

--- --- --- ---

Insert Figure 3 Here.

--- --- --- ---

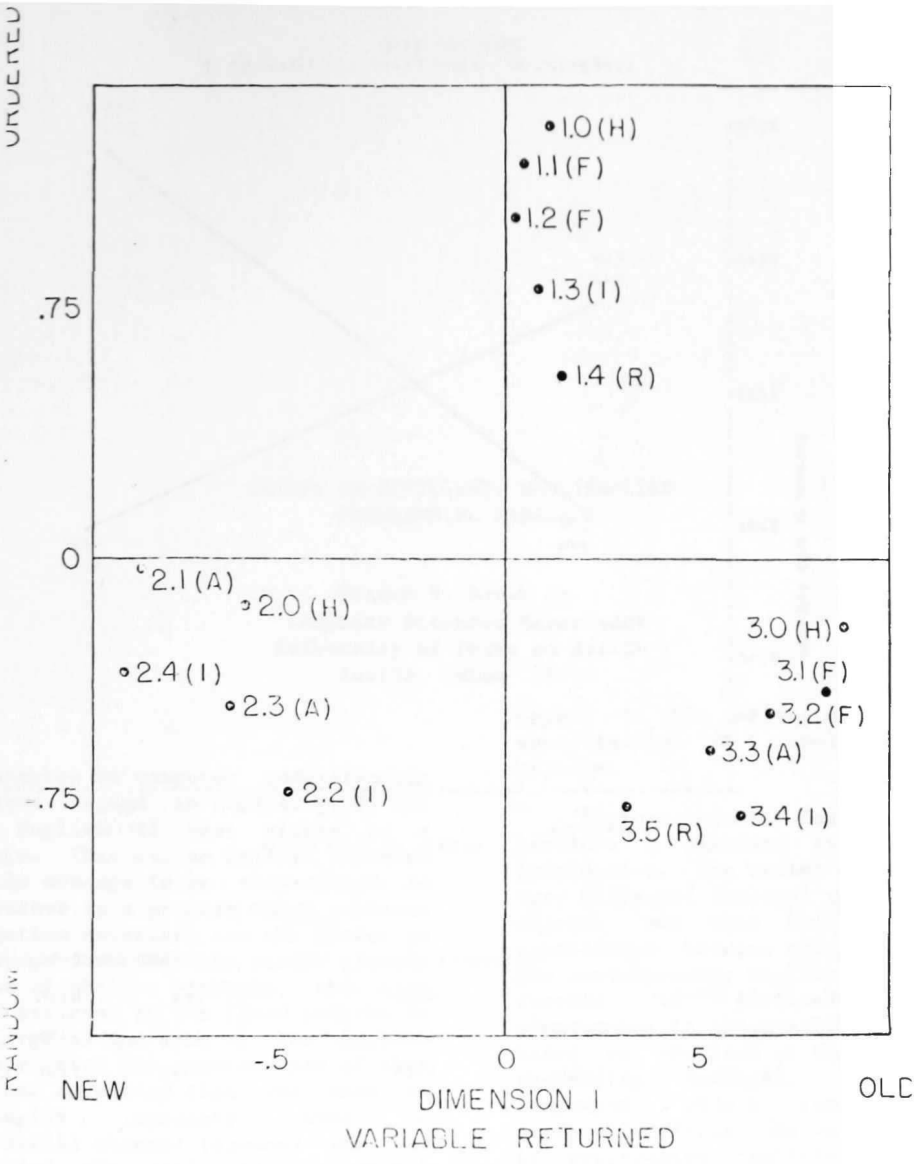
Looking at the left hand side of Figure 3 we see that the Novices are actually faster than the Experts when answering a High Detail question although we see on the right hand side that the Experts are faster than the Novices when answering a Low Detail question. That the Novices are faster than the Experts when having to answer a High Detail question indicates that the Novices are used to forming High Detail, that is non-abstract, specific representations of how programs function. That the Experts are faster than the Novices when having to answer a Low Detail question indicates that the Experts are used to forming Low Detail, that is abstract representations of what a program is doing. The error rates for the non-congruent and Code Only conditions both show the same interaction, Novices do well on High Detail questions while Experts do well on Low Detail questions.

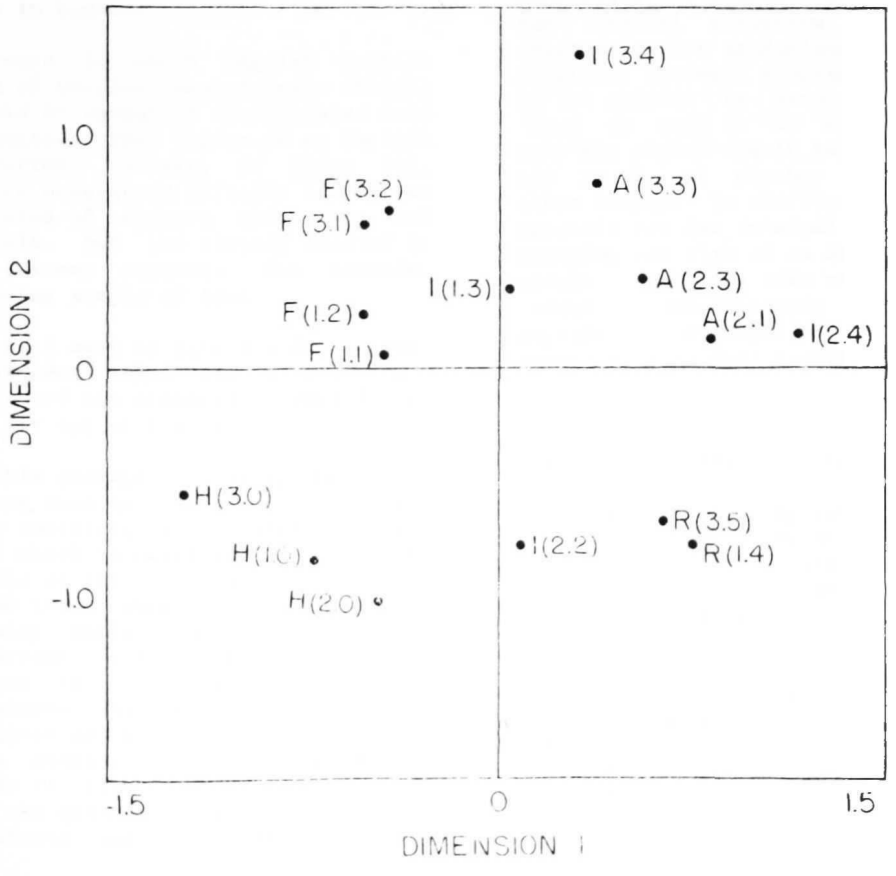
B. Congruent Conditons (here the level of detail of the flow chart and the question match)

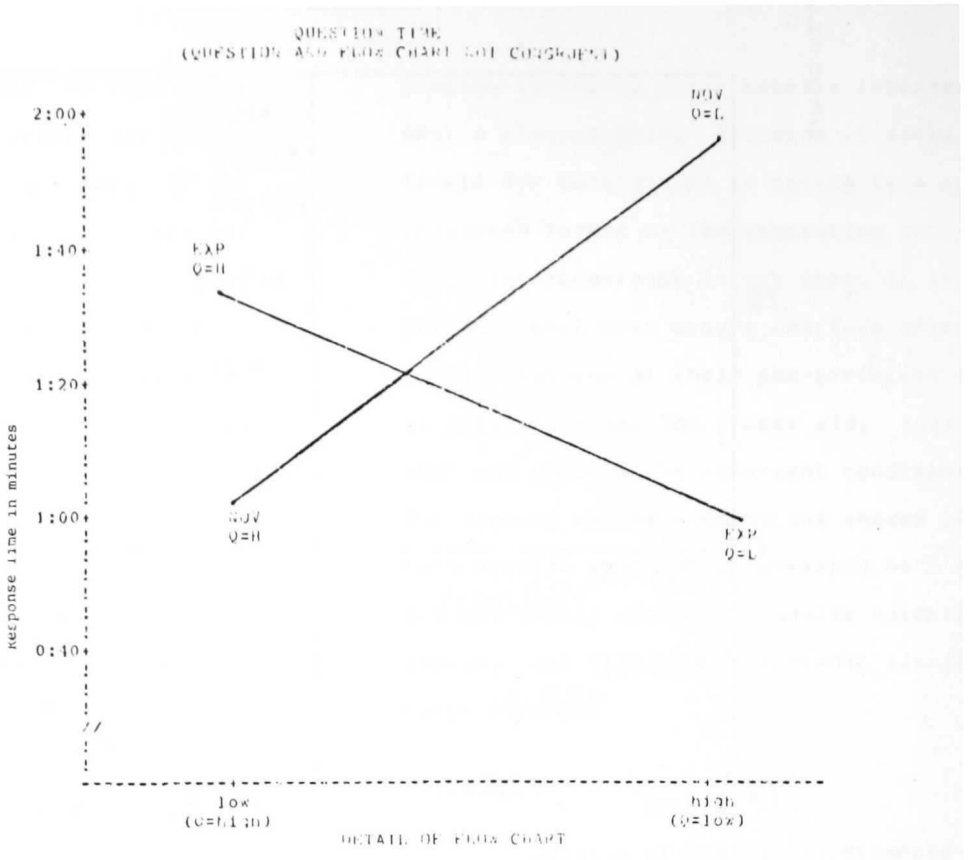
The results of the Non-congruent Conditions suggest that Novices do represent how a program functions while Experts represent what a program does. Although it seems difficult for both groups to switch from their preferred levels of representation when they have been encouraged to use them, it is still possible that both groups can form effective representations at their non-preferred levels if they are given the proper aid. This is what was done in the congruent conditons. The results of these conditons showed that both Novices and Experts answered both High and Low Detail questions equally quickly suggesting that they can be somewhat flexible in their encoding.

Conclusions

The results of the two experiments reported here suggest that Novice and Expert computer programmers represent information about programs differently. The representations of the Experts is more abstract and is based on what the program does, while the representation of the Novices is more detailed and is based on how the program functions. These differences also influence the utilization of the information being represented.


ORDERED
1.0 (H)
1.1 (F)
1.2 (F)
1.3 (I)
1.4 (R)
.75
0
2.1 (A)
2.0 (H)
2.4 (I)
2.3 (A)
2.2 (I)
-.75
3.0 (H)
3.1 (F)
3.2 (F)
3.3 (A)
3.4 (I)
3.5 (R)
RANDOM
-.5
0
5
NEW
DIMENSION 1
OLD
VARIABLE RETURNED


1.0
I (3.4)
F (3.2)
A (3.3)
F (3.1)
I (1.3)
A (2.3)
F (1.2)
A (2.1)
I (2.4)
F (1.1)
0
DIMENSION 2
H (3.0)
R (3.5)
H (1.0)
I (2.2)
R (1.4)
-1.0
H (2.0)
-1.5
0
1.5
DIMENSION 1



MEANS

	CONGRUENT		NOT CONGRUENT	
FLOW CHART =	LOW	HIGH	LOW	HIGH
NOVICE	1:14	1:22	1:01	1:55
EXPERT	1:03	1:10	1:32	1:01

GLISP: AN EFFICIENT, ENGLISH-LIKE PROGRAMMING LANGUAGE

Gordon S. Novak Jr.
Computer Sciences Department
University of Texas at Austin
Austin, Texas 78712

1.0 INTRODUCTION

My earlier research on computer understanding of physics problems stated in English [1,2] has convinced me that English is best viewed as a programming language. That is, an English sentence does not contain the message to be transmitted to the reader, but rather is a program which provides the minimum information necessary for the reader to construct the message from what the reader already knows. In the case of physics problems, the size of the model constructed by the ISAAC program is some 30 times as large as the size of the English sentences which specified the construction of that model. Woods [3] has suggested that the need to communicate complex concepts over a bandwidth-limited serial channel (speech) was the driving force behind the evolution of natural language abilities in humans.

Study of the ways in which English permits compact expression of complex ideas reveals several features which would be useful if incorporated into programming languages. The reader of an English text maintains a current context, or focus [4], which can be used to understand definite references to objects or features of objects which are not specified completely, but are closely related to objects in the current context. For example, consider the following sample of text:

> Last night I went to Scholz's for a beer. The bartender asked for a ride home, since his car was disabled. Somebody had let the air out of the tires.

A person reading this passage can easily understand a definite reference such as "the air", which means "the air which was contained in the tires which are part of the car which is owned by the person who works as a bartender at the bar named Scholz's". The reader has made these connections while reading the story by using world knowledge and by maintaining a current context relative to which definite references to previously unmentioned objects and features can be understood; each reference to an object or feature causes it to be brought into the context, thus enabling further references relative to it. The ability of the reader to infer the connection between a definite reference in a sentence and a "closely related" object in the current context permits the compact specification of complex relationships among objects.

Another valuable feature of English is that it provides a standard interface for communicating information. The writer and the reader may have very different internal representations for certain objects, but they both have procedures for translation between their internal representations and corresponding English descriptions. A related feature is provided by "object-oriented" programming in the SMALLTALK language [5], which is based on the idea of Objects which communicate by exchanging Messages. In most programming languages, object representations are merely storage locations; the nature of the representation is represented implicitly in the programs which manipulate the storage locations. In SMALLTALK, the internal structure of objects is hidden, and programs cannot manipulate the internal structures directly; instead, programs query and change values in the objects by sending them messages, e.g., "what is your X?" or "set your X to the value y". Only the object itself knows whether it actually has an X, or whether its X is a consequence of other values. In addition, the object can act to maintain its own internal consistency; for example, changing the size of an object may require that the object change the size of its picture on a display screen. Unfortunately, SMALLTALK has been implemented on special hardware, and has been unavailable to most researchers.

2.0 NEED FOR ENGLISH-LIKE PROGRAMMING LANGUAGES

English-like programming provides two features which are needed by workers in Cognitive Science and Artificial Intelligence and which are not provided by most existing programming languages: brevity of expression and ease of changing representations. In fact, these two are intertwined: the more detail one has provided about how to perform an action on an object, the more code one will have to change if the basic structure of the object is to be changed. Most existing programming languages implicitly specify the structures of objects within the code. For example, in either PASCAL or CLISP [6], referencing

a field of a record structure requires that both the record and a complete path from the record to the desired field be specified in the code; if the record structure is to be changed, all the code which references such records will often have to be changed also. In a large system, such changes are so difficult that significant changes to data structures are seldom possible once a large body of code exists.

3.0 GLISP

GLISP is a LISP-based language which permits English-like programs containing definite references. GLISP is implemented by a compiler which compiles GLISP programs into LISP relative to a knowledge base which is separate from the programs; the resulting LISP code can be further compiled to machine language by the LISP compiler. In GLISP, the execution of a program causes an implicit context of computation to be constructed, just as an English conversation causes an implicit conversational context to be constructed in the minds of the conversants. The context is computed at compile time, using flow analysis, from the previous context, which includes Structure Descriptions of previously mentioned objects. Definite references to features of objects which are currently in context are permitted; these cause the newly referenced objects to be added to the context, allowing further references relative to them.

The initial context within a GLISP function consists of the arguments of the function, its PROG variables, and any declared global variables. The context contains, for each variable, its variable name, reference name, and Structure Description. When a definite reference is encountered within a GLISP program, the compiler determines whether the reference names such a variable or names a substructure or feature of some variable which is in context. If a substructure or feature is referenced, the compiler determines how to get it from the original structure; the resulting code replaces the definite reference in the compiled version of the program. In addition to producing code to get the feature from the starting structure, the compiler also determines the Structure Description of the result. The new item and its Structure Description are added to the context, thus enabling further definite references relative to it. When the compilation is finished, the context structures disappear; the compiled code contains only the LISP code necessary to perform the specified actions. Thus, the code produced by GLISP is relatively efficient; the user of GLISP must pay for compilation, but does not incur a runtime penalty. The GLISP compiler runs incrementally, so that functions are compiled automatically the first time they are called.

The following example illustrates some of the features of GLISP. Suppose that a wicked witch curses a grandmother by decreeing that each of her calico cats shall age by five years. The code to accomplish this can be written in GLISP as follows:

```
(CURSE (GLAMBDA ( (A GRANDMOTHER) ) (PROG ()
   (FOR EACH CAT WITH COLOR = 'CALICO
        DO  AGE ←+ 5)   )))
```

The GLAMBDA indicates that this is a GLISP function, and causes the GLISP compiler to be called when the function is first interpreted (using the LAMBDATRAN feature of INTERLISP [6]). Since GLISP maintains a context and permits definite reference, it is often unnecessary to give names to variables; thus, we need only declare the type of the argument, (A GRANDMOTHER). Since a GRANDMOTHER is in context, the compiler can determine how to access her CATs and how to generate an appropriate loop to examine each of them. Within the loop, of course, the current CAT is in context, allowing definite reference to its features. The compiler generates the appropriate kind of test to compare the COLOR of the CAT against the constant 'CALICO; if needed, the constant and the operator could be coerced into the appropriate forms. For example, 'CALICO might have several possible meanings; in the context of the COLOR of a CAT, it could be coerced to the unique constant 'CALICO-CAT-COLOR. If the test is satisfied, the AGE of the CAT is increased by 5; the operator ←+, which specifies appending when applied to lists, is interpreted as addition when applied to numbers.

In the GLISP program, we have implied that certain objects have certain features, e.g., that a CAT has a COLOR, but we have said nothing about how to get or replace the COLOR of a CAT, or about what type of entity the COLOR actually is. This information is held separately in the knowledge base of Structure Descriptions and other information relative to which the program is compiled. This makes possible significant changes to data structures with no changes to the code -- a goal long sought in high-level languages, but one which has been largely unrealized for structures involving pointers. GLISP can be viewed as similar to SMALLTALK in the sense that a program does not

specify directly how to manipulate objects. Instead of sending a message to an object, we can think of the GLISP compiler as generating the code to do what the object would do if it received such a message. This provides some of the flexibility of SMALLTALK with high runtime efficiency.

The GLISP compiler allows GLISP expressions and ordinary LISP to be mixed; the user can use as much or as little GLISP as desired. A Structure Description language is provided for the common LISP data structures, and the compiler automatically generates code to access such structures. In addition, the compiler provides a clean interface to one or more representation languages; the user can use both ordinary LISP structures and units in his favorite representation language, accessing both in a transparent manner. A more compact, CLISP-like syntax for GLISP expressions is provided in addition to the English-like syntax. The GLISP compiler, accessing both LISP structures and our GIRL representation language [7], is currently running.

4.0 ACKNOWLEDGMENT

This research was supported by NSF Award No. SED-7912803 in the Joint National Institute of Education - National Science Foundation Program of Research on Cognitive Processes and the Structure of Knowledge in Science and Mathematics.

5.0 REFERENCES

1. Novak, G., "Computer Understanding of Physics Problems Stated in Natural Language", American Journal of Computational Linguistics, Microfiche 53, 1976.

2. Novak, G., "Representations of Knowledge in a Program for Solving Physics Problems", Proc. 5th IJCAI, Cambridge, Mass., 1977, pp. 286-291.

3. Woods, W. A., Symposium on Formal Semantics and Natural Language Processing, University of Texas at Austin, March, 1979.

4. Grosz, B., "The Representation and Use of Focus in Dialogue Understanding", Ph.D. thesis, University of California, Berkeley, 1977. Also Technical Note No. 151, SRI International, Menlo Park, California.

5. Ingalls, D., "the Smalltalk-76 Programming System: Design and Implementation", 5th ACM Symposium on Principles of Programming Languages, Tucson, Arizona, January 1978.

6. Teitelman, W., "INTERLISP Reference Manual", Xerox Palo Alto Research Center, 1978.

7. Novak, G., "GIRL and GLISP: An Efficient Representation Language", submitted to IJCAI-81.

SOME QUESTIONS ABOUT VERBAL AND PICTORIAL REPRESENTATION

Sheldon Richmond
Box 1443, Station B
Downsview, Ontario

Parallels between distinct regions are intriguing. They suggest previously unsuspected unities. For instance, the parallels among the laws of radiation, magnetism, and electricity. Of course, we may impose a parallel upon two regions suggesting an unreal unity. For instance, paralleling laws of historical development and human social organization along the lines of biological evolution suggests an unreal unity between culture and biology. Thus, parallels suggest unities, but the unities should always be approached with a questioning-attitude. It is with this attitude that I wish to approach the parallel hypotheses presented by E.D. Hirsch,Jr., and E.H. Gombrich about the history and psychology of verbal and pictorial representation, respectively. The parallel aspect of these hypotheses raise important questions about cognitive processes connected with verbal and pictorial structures: Is there one cognitive process underlying reading words and seeing pictures? Do reading and seeing pictures appear to the mind in the same way?

If we synthesize the work of Hirsch and Gombrich, we arrive at yes-answers to both questions. This result is surprising. It counters the common assumption that verbal and pictorial representation are very different modes of cognition. The old saying "a picture is worth a 1,000 words" may be literally true. Pictures may be dense versions of written verbal representations. So, thinking pictorially and thinking verbally may be one mental process. This, then, is the purpose of my paper: to raise questions about a possible unity between picturing and verbalizing suggested by parallels in the work of Gombrich and Hirsch.

1. Hirsch on Composition:

Hirsch asks: are there universal rules for composition? His answer involves developing a theory of the psychology of reading and applying this theory to the history of composition and the process of reading. The history of written language progresses toward increading readability. The more easily a passage can be read, the more readable it is. The less time one can see through a passage, as it were, the more easy it is to read. Thus, principles of style--developed by a process of trial-and-error throughout the history of writing--tend towards 'economy', 'simplicity', 'variety within unity'... for the purpose of conveying the author's thought (including intentionally obscure and complex thought). In short: "...relative readability is an intrinsic and truly universal norm of writing".[1]

This thesis of the universality of the goal of increading readability is intertwined with the hypothesis of "linguistic universals". The mind processes all written texts in the same way. For instance, the mind looks for patterns that when not found lead to uncertainty; and when too often found lead to inattention:

> To avoid wasteful attention shifts, the verbal theme of one phrase must be similar to the theme of the preceding phrase. The expectancies set up by one phrase should be fulfilled in the next...[2]

At this point, the reader may wonder how verbal patterns appear in the mind? Do they appear as word-like? Hirsch has a provoking answer. The short-term memory holds verbal patterns almost literally. Consequently, the short-term memory which comes into immediate contact with the page, looks for similarity in verbal patterns for ease of retention. However, the long-term memory contains semantic contents and recalls the meanings through seeking thematic representations or labels that surround the meanings. From this difference in the ways short and long-term memory process verbal patterns, two consequences occur for the reading of texts: 1)The short-term memory best operates on clauses--discrete units of verbal patterns that recurr in the text.

> Readability is enhanced when closure is rapid and stable, since rapid and stable closure greatly reduces both processing time and the burden on short-term memory.[3]

2)The long-term memory is stimulated by labels or "thematic tags" that refer to a variety of specific ways of verbalizing meanings.

> Since the meaning of the whole discourse (remembered and expected) is mainly stored in a nontemporal, nonlinguistic form, the writer will assist the reader by continually repeating a rather small number of thematic tags which represent that remembered (and expected) holistic meaning...[4]

In short, the process of reading involves the application of nonlinguistically stored semantic contents to verbal patterns that require repetition for storage in the short-term memory. Without enough repetition of verbal patterns, the short-term memory cannot grasp the text in order to transfer the semantic contents to the long-term memory. However, with too much repetition, the mind loses attention. Likewise, without the use of labels, semantic contents cannot be realled for application to specific verbal patterns; but with the use of too many labels too often, semantic contents become dulled.

Hirsch's theory of readability may be extrapolated to a theory of how thoughts appear in the mind: how does the mind's eye see thoughts? We think with nonlinguistic semantic contents. The mind's eye sees nothing, but the mind's 'hands' shape semantic contents. Words stimulate and simulate nonlinguistic thinking. For instance, readability depends upon striking a balance between the use of thematic tags and specific, recurrent, verbal patterns: thematic tags set up a context of expectations, and recurrent verbal patterns fulfil and modify those expectations. This way of structuring texts simulates the process of specifying and modifying general meanings: our thinking involves the formation of general themes that are applied to specific situations--and so, thinking in 'themata' involves setting up expectations that when unsatisfied leads to the refashioning of those 'themata'. Thus, the more readable a text is the more recognizable is its illusion of thought-processes: the formation and refashioning of semantic worlds or themata.

In short, my extrapolation of Hirsch's theory of readability to cognitive processes is this: thinking may be a nonlinguistic process of which verbal patterns provide an illusion; the greater the illusion, the more readable the text--the more transparent the meaning or thought. Though we don't think in words, we think with words: words, by simulating cognitive processes, provide a track for directing and improving our cognitive processes.

I will come back to this theory of the illusory nature of verbal representation after turning to Gombrich's parallel discussion of the representative nature of pictorial illusion.

2. Gombrich on Pictorial Representation:

Most people assume that looking at real objects is radically different from looking at pictures. Gombrich has taught us that this natural way of thinking about pictures is mistaken. The mental processes involved in looking at pictures and objects are the same: both involve expectations, corroborations, and counter-expectations. This surprising theory of the cognitive processes involved in perception provokes the question: how do we ever tell the difference betweeen pictures and objects? The answer is obvious: when you try to put your head through a painting of an open window, the canvass

or wall gets in the way:

> ...we still experience some kind of illusion when we see a picture on a wall or in a book --from a point, that is, where the perspective should go wrong. Here as always we first read the picture for consistency, and this consistency, the interaction of clues, is not wholly upset by our changing viewpoint. The painting may cease to be consistent with the world around it, but it reamins closely knit within its own system of references.[5]

I want to explain Gombrich's theory in some detail--in order to bring out its parallels with Hirsch's theory. In seeing pictures, according to Gombrich, our minds match the conventionalized modes of representing reality against the picture; and match the picture against reality. Our minds test the degree of recognizability of the picture. Moreover, the history of pictorial representation in art is a history of making and matching. The cognitive process of making and matching has an overflow into the nature of veridical perception: we find that veridical perception also involves making and matching. Just as in painting pictures, artists use conventional schemata as hypotheses about how we see reality, so too in attempting to see objects and events in the real world, our mind employs hypothetical schemata that are either corroborated or refuted by reality:

> It is the power of expectation rather than the power of conceptual knowledge that molds what we see in life no less than in art...Every time we scan the distance we somehow compare our expectation, our projection, with the incoming message. ...Here as always it remains our task to keep our guesses flexible, to revise them if reality appears to contradict, and to try again for a hypothesis that might fit the data.[6]

By now the thesis of the inferential nature of perception is well-known. However, this thesis raises an important but unaksed question about the nature of mental representations: how does the mind's eye see the correction of false visual hypotheses? Does the mind see pictures that are redrawn?

For theoretical reasons the answer is no: the mind's eye sees nothing; rather, the mind's hands reshape nonvisualized hypotheses about the world. Given Gombrich's theory that pictorial representation has a history because the mind makes and matches visual schemata against reality, it follows that the visual schemata teach us how to see:

> ...The wish to find confirmation of some new experiment may make the progressive suggestible and may thus facilitate the artist's task of modifying his code...we genuinely recognize pictorial effects in the world around us, rather than the familiar sights of the world in pictures.[7]

The fact that pictures sometimes teach us how to see, falsifies the commonly held thesis that in painting we copy our internal three-dimensional mental pictures onto a two-dimensional plane. No recognition of reality could be possible if we were only copying our mental images when we paint--we could only see reflections of our minds rather than reflections of reality (and reflections of art in reality). Since painted images of reality both deceive and inform the eye[8], we learn how to see reality by way of testing our pictorial representations and nonvisualized internal (mental) schemata informing our visual expectations against reality. When 'introspecting' upon the process of correcting mental schemata, though we 'see' nothing, we feel the impact of our mind's hands reshaping those nonvisualized schemata that reset our expectations about visual reality.

At this point, one may wonder: do we ever have mental images? Taking off from Gombrich's theory of making pictorial schemata and matching them against reality for recognizability, I hypothesize that we learn to have mental images by introjecting visualized pictorial schemata. The mind's eye is a construct derived from seeing representational pictures. Consequently, there is no natural way of seeing reality: no fixed repertoire of mental images that we attempt to impose upon reality. Rather, we learn to see reality in specific ways through the construction and testing of visualized hypotheses or pictures. We learn to have different forms of mental images by introjecting different schemata.

In short, if we could construct a camera for taking pictures of mental images, we would only see copies of introjected pictures. When the mind is actively seeing--probing reality--it employs nonvisualized projections of reality: the mind 'feels' reality with 'mental shapes' that are refashioned by the way of the process of making and matching.

3. Parallels between Reading and Seeing-Pictures:

For the sake of brevity, I tabulate the parallels between Hirsch and Gombrich:

Questions/Answers	Hirsch	Gombrich
How does the history of ____ progress?	composition	pictures
By making and matching schemata for_____.	readability,	recognizability.
How do we read see———	texts?	pictures?
By testing anticipations against the _____.	text.	picture.
-----Extrapolations----		
How do the mental contents of _____appear in the mind?	verbal/	pictorial/ schemata
They appear as____contents (schemata).	non-linguistic	non-visualized
How does the mind correct its internal schemata?		
By matching its schemata against introjected_____.	verbal/	pictorial/ representations

These parallel questions and theses prompt the idea of a single process informing the two distinct areas.

4. Is there a common psychological process for reading and seeing-pictures?:

This question, suggested by the parallels between Hirsch and Gombrich, involves searching for a single mental process for the representation of reality through words and through pictures. Though both forms of representation--words and pictures--have different manners of referring to reality, we wonder whether they are grasped and used by the mind in one way with one mental operation. For instance, according to Nelson Goodman[9], though pictures and words are non-notational (i.e. unlike musical scores where each note on a scale stands for a specific sound), but words not pictures are syntactically and semantically differentiated. In plain words, words are distinct and refer to distinct regions of reality; but pictures contain forms that merge and refer to overlapping areas of reality. However, the question we ask here seeks to uncover a common mental process for grasping the very different symbolic systems of words and pictures. How could this common mental process work, if it were to exist?

The cognitive contents of mind are nonsymbolic and nonrepresentative. Our minds shape its contents by simultaneously introjecting and retrojecting verbal or pictorial patterns. Our minds project shapes upon reality that are perceptually refined by the use of introjected pictorial schemata and counter-expected verbally coded perceptions. The flexible mental fields of our minds take on increasing representative content through using mismatches of introjected verbal/pictorial schemata, and mismatches of reality, to reshape the mental fields.

In sum: synthesizing the theories of Hirsch and Gombrich permits us to seek for an unified mental process for thinking and seeing. The process may be this: We probe reality with simulations of thinking--words--and simulations of reality-seen--pictures. These probes occur in the mind as projected nonsymbolic fields and appear to the mind's eye as introjected verbal patterns and pictorial schemes. It is easy to confuse the introjected contents of our mind with the actual mental processes of thinking/seeing: the process of holding and shaping contents with our mind's 'hands'.

NOTES

1. p.89,E.D.Hirsch,Jr., The Philosophy of Composition (Chicago, University of Chicago Press, 1977)
2. p. 107, ibid.
3. p. 119, ibid.
4. pp. 123-124, ibid.
5. p.227, E.H.Gombrich, Art and Illusion(New Jersey, Princeton University press,1961).
6. p.225, ibid.
7. p.237, Gombrich,"Visual Discovery through Art", ed. James Hogg, Psychology and the Visual Arts (Harmondsworth, Penguin, 1969)215-238.
8. See p.314, Gombrich, op.cit.
9. Nelson Goodman, Languages of Art (Indianapolis, Bobbs-Merrill,1968); Sheldon Richmond,"A Discussion of Some Theories of Pictorial Representation", Dialectica,34,3(1980)229-240.

The relationship between human motion and objects in the cognitive representation of visual events.

Margot D. Lasher
Assumption College

This paper concerns the cognitive representation of events which involve human motion.* I have elsewhere proposed a schematic structure for events which involve only the motion of the body, as in some athletics and dance (Lasher, 1981). In this paper I would like to describe the relationship between the motion of the body and objects which may be part of the event. The eventual goal is to develop schematic representations for all types of events of human motion.

The paper will look at two pairs of event units. The first event of each pair involves only the motion of the body; the second involves that same category of motion in relation to an object. Certain aspects of these paired events can be identified without further analysis. We are dealing with only one type of event unit, an event involving voluntary human motion. The AGENT of the event will always be the person in motion. The ACTION will consist of a description of that person's motions.

Event 1: A person leans, reaches out, and brings the arm back toward the body.

Event 2: A person leans, reaches out, picks up a glass and brings the glass back toward the body.

Figure 1 provides schematic structures for Events 1 and 2. Event 1 describes a motion unrelated to any object. It might be a movement in athletics or dance. The ACTION consists of a preparatory motion and a completing motion. There is experimental evidence for the psychological coherence of this preparatory-completing structure in the perception of events involving human motion (Lasher, 1981). A psychologically coherent event unit consists of a small, unfixed number of preparatory motions followed by a completing motion. Preparatory motions tend to be relatively stable motions, while completing motions are relatively unstable motions. The unstable completing motion is encoded as the intention of the entire event unit. The event unit is encoded as completed when this relatively unstable motion is finished and the person has returned to a position of stability in relation to the ground.

Each motion, whether preparatory or completing, can be described in terms of both the changes in the human body itself, labeled MOTION in Figure 1, and in terms of the variables which influence that motion, labeled MOTION RELATIONS. There are a potentially infinite number of motion relations. The actual representation of any event in a real mind will depend upon what variables are attended to at the time. I have assumed, for purposes of illustration, that we are attending to the direction of the lean and reach, which is forward, and to nothing else.

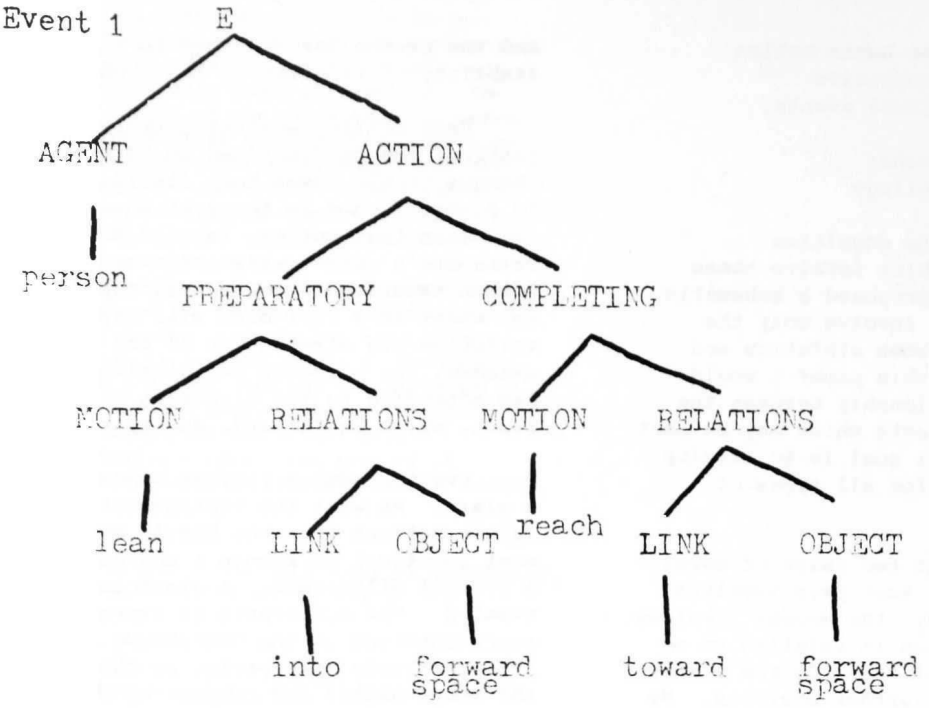
Event 2 adds a visible object to the scene, a glass. We want the representations to accurately reflect that the bodily motions are almost identical in Events 1 and 2, yet there is a crucial difference: a glass is picked up in Event 2. The difference is expressed in the representation of the COMPLETING MOTION, which is cognitively interpreted as the intention of the event unit. The completing motion is a motion which contains as one of its internal positions that of the fingers closing in upon and pressing against the glass. I have arbitrarily used the words reaches-picks up to represent this motion. The fact that in language these are two separate verbs is irrelevant to the representation of the visual event unit.

The MOTION RELATION which most influences the form of this completing motion is the OBJECT, the glass. The verbal nature of the representation is again not relevant. Holding might seem to be a separate motion because it is a verb rather than a preposition. From the viewpoint of the visual event, it is equivalent to a preposition: it is a link between the bodily motion and some other spatially represented variable which influences the form of the body's motion. The holding is not a separate motion. The body returns to a stable, balanced position, with the arm near the central axis of the torso, at the ends of both paired events. In Event 2, however, the body returns with a glass.

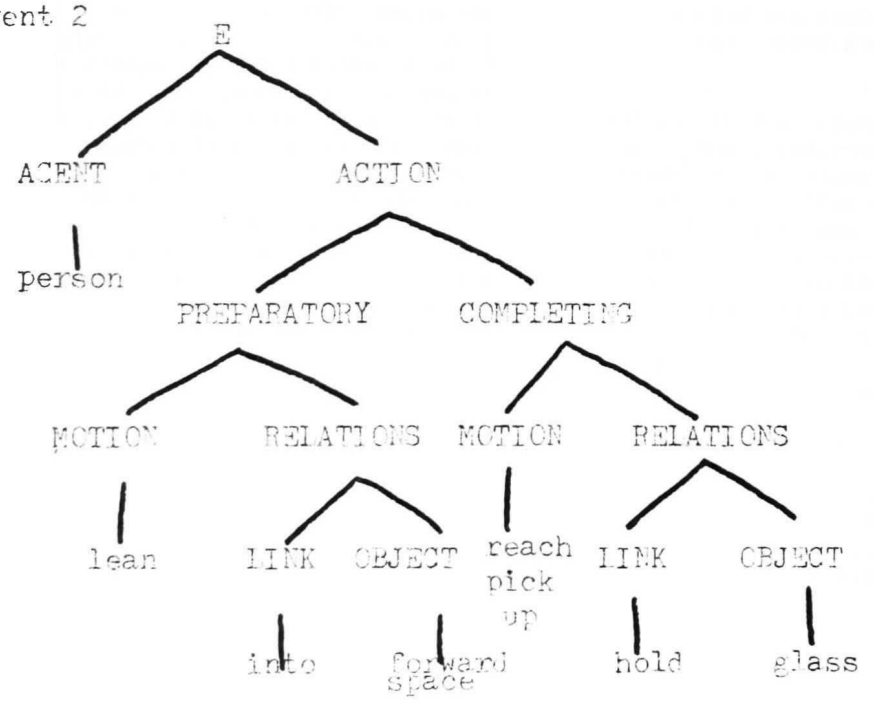
Events 1 and 2 were originally paired because they both belong to a general category of human motion events, reaching motions. Laban has spoken of such abstract categories of motion (Laban, 1966). We reach to pick something up, to hang something up, to put something down. The motion in bowling and the pitch in baseball are full reaching motions. When the reaching motion is extended to its fullest without the presence of an object, and one leg is extended

Figure 1

a. Event 1



b. Event 2



backwards in order to balance the reaching arm, we have an arabesque in ballet.

An interesting thing about human motion events is that any event involving voluntary human motion can be described in terms of AGENT-ACTION, without an OBJECT: any ACTION which can be performed in relation to an OBJECT can also be performed without the OBJECT. But in terms of a description of the visual structure, there is always a spatial direction involved. The person reaches upward or downward, toward the sky or toward the ground. In training for athletics or dance, aspects of the environment that are taken for granted in ordinary motion are often brought to conscious attention: the athlete is trained to notice the relationship between the body and the ground, and the body and the space around it. If the spatial aspects of the event are considered in relation to the motion itself, these spatial aspects act in the same ways as an object would act: as variables relevant to the intention of the motion.

Because of this similarity between visual directions and ordinary objects in a scene, I have represented them equivalently under the MOTION RELATIONS node. The variable is represented as the end-point of the motion under the OBJECT node. The LINK node is required to represent exactly how the OBJECT influences the MOTION. In Event 1, for example, the preparatory motion of leaning actually goes into, or bodily fills, the forward space; the completing motion of reaching goes toward, but not into, the forward space. Of course the forward space of the preparatory motion can be a different physical area from the forward space of the completing motion: space is forward or backward only in relation to the body.

The addition of an object in the second pair of events takes us into the interesting problem of representing causal relationships between event units.

Event 3: A dancer steps and leaps.

Event 4: A basketball player steps and takes a jump shot.

Figure 2 shows the visual relationships between the motion of the person, and the motions of the person plus the ball, in Events 3 and 4. Schematic representations of these events are given in Figure 3.

Until point 5 (Fig. 2) the visual structure of Events 3 and 4 overlap. At point 6, the person begins to descend to the ground (Events 3 and 4) and the ball continues to rise (point 7) until it falls at point 8 (Event 4 only). In both Events 3 and 4 the person must descend to the ground. But the ball has it own, independend motion in Event 4. We have in Event 4 two separate motion descriptions to deal with: the person in motion, whose structural description is almost identical to that of Event 3, and the object in motion, which spatially and temporally goes through its own motion.

We want our representations to mirror the similarity of the person's motions in Events 3 and 4. We also want the representation of Event 4 to correspond to the intuition that the basketball player's jump shot, and the basketball's descent through the hoop, are parts of the same event. We do not want to be in the position of saying that Event 4 is represented by two loosely joined schematic structures, one representing the players motion and the other the motions of the ball. These considerations lead to schematic structures which are very similar for Events 3 and 4, and in which the motions of the ball in Event 4 are part of the higher event of the person's motions.

Nevertheless, there is a real difference between the ball in Event 4 and the object in Event 2 (the glass which is picked up): the ball has a motion which is spatially and temporally non-co-extensive with the motion of the agent. In a basketball game people tend to watch the motion of the ball, and for many people, the event ends when the ball goes into or misses the basket, not when the player has descended again to the ground. Attaching the ball as a MOTION RELATION to the ACTION of taking a jump shot, as the glass is a MOTION RELATION to the ACTION of reaching/picking up, captures only part of the experience.

In order to represent these relationships I have described Event 4 as an event unit containing an embedded event unit (Figure 3b). As long as an object is carried, pushed, or pulled by a person, so that the object's motion retains a one-to-one relationship with those of the person, the object's motion is not represented as a separate unit (ie: the glass in Event 2). When the object's motion ceases to have a one-to-one correspondence with those of the person, the object's motion becomes represented as a separate event. However, the object's motion in this case is a consequence of the person's; the ball goes into the basket because the player steps, jumps, and releases the ball in a certain way. Players and coaches can often predict at the moment the ball leaves the player's hands whether the ball will hit the basket. This de-

Figure 2: Events 3 and 4

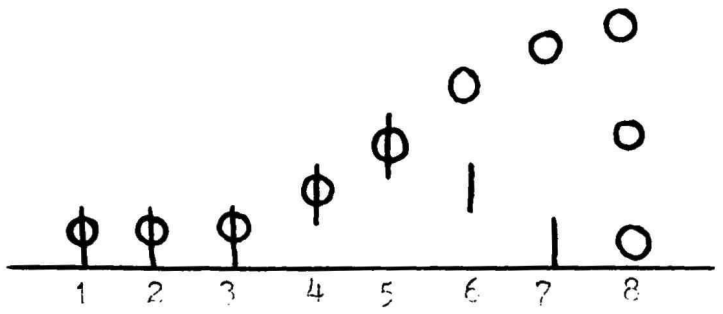


Figure 3
a. Event 3

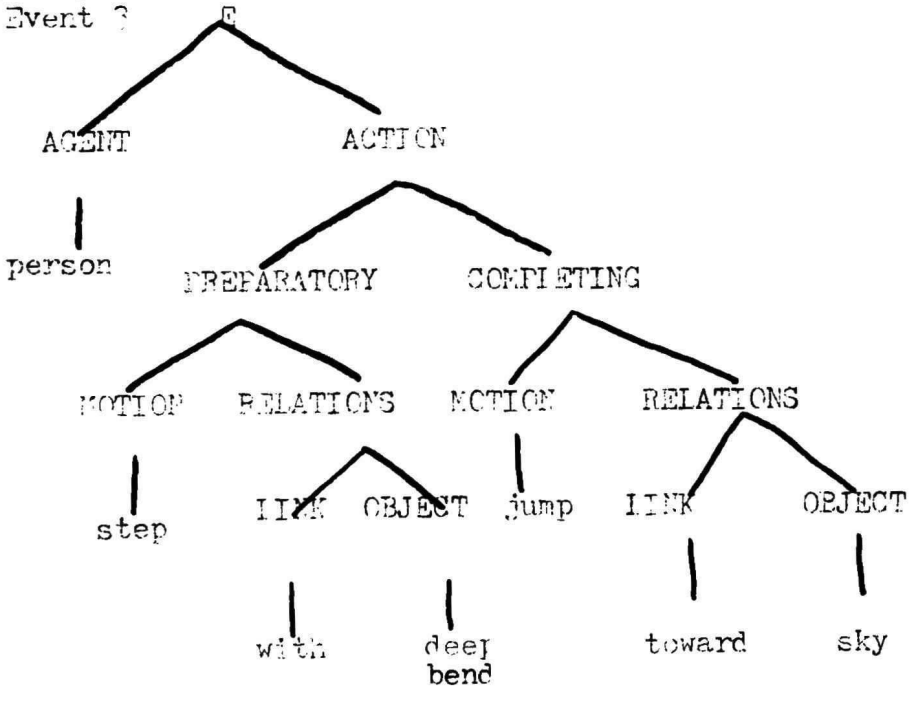
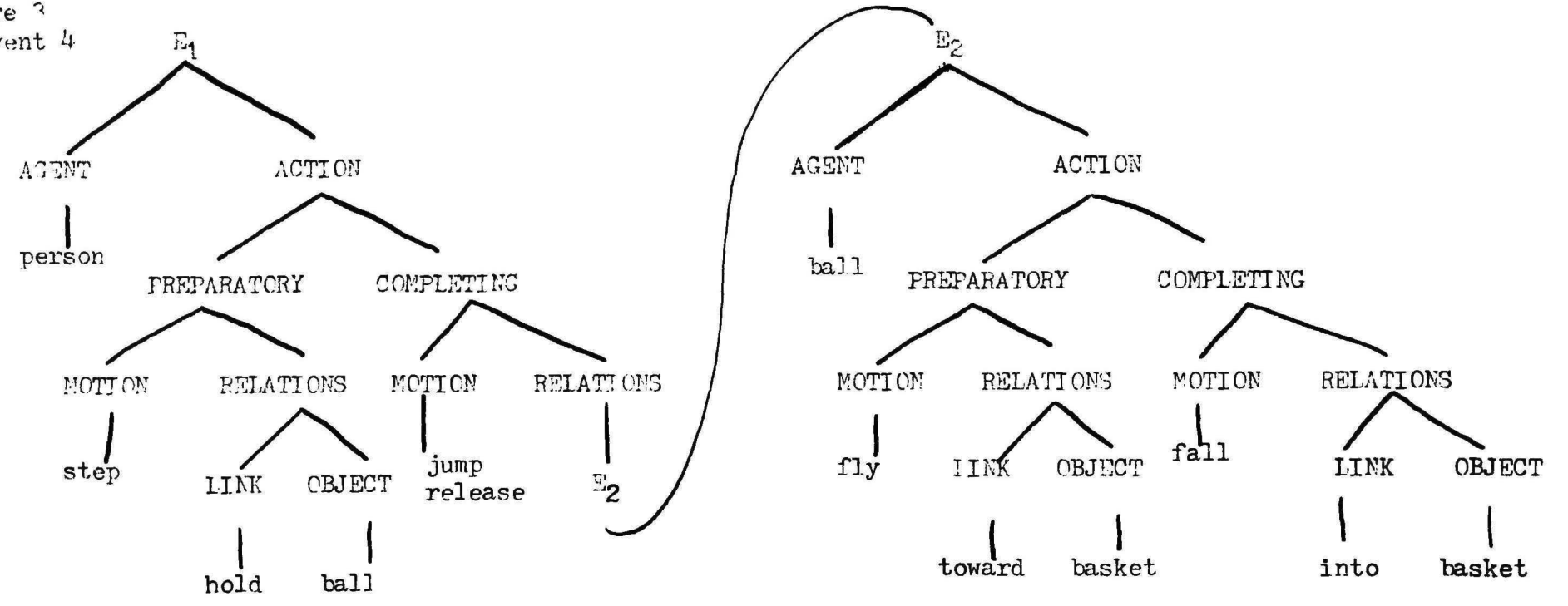


Figure 3
b. Event 4



pendence of the ball's motion upon the player's motion is captured by the embedded structure.

Events 3 and 4 fall into another abstract category of human motion events, run-leap. This category is especially interesting because the body defies gravity in a visually dramatic way at the center of the completing motion. When we are watching just a person in motion we normally watch the person descend to the ground at the end of the leap. The event is fully completed only when the person has returned to a position of stability on the ground.

But in the case of an embedded event in which the object has an independent motion of its own, the eye tends to follow the motion of the object (the ball, for example) once it leaves the person's hands. The follow-through aspect of the completing motion is not actually perceived. The dunk shot in basketball, however, is an interesting exception. In the dunk shot, the player's and the ball's descent occur together temporally and are very close together spatially. It is an interesting possibility that the excitement of the dunk-shot for spectators is partially due to the synchrony of endings of the embedded and embedding event units (Phillippe Poisson, personal communication).

Our excitement may arise from our ability, in this particular event, to actually see both the player and the ball return to the ground. This excitement would really be an aesthetic kind of excitement: the experience of a perfect fit between the schematic cognitive structure and the particular manifestation of that structure being observed.

References

Laban, Rudolf. The Language of Movement: Choreutics. London: Macdonald & Evans, 1966.

Lasher, Margot D. "The cognitive representation of an event involving human motion." Cognitive Psychology, 1981, 13, in press.

* I would like to thank John M. Carroll, Joseph M. O'Brien, and Philippe L. Poisson for helpful discussions of the ideas in this paper.

Toward a Model of Cognitive Process in Cartoon Comprehension

Michael E. J. Masson and Inga Boehler
University of Victoria, Victoria, Canada

The comprehension of humorous material involves an important interaction between use of general world knowledge and the encoding of new verbal or pictorial information. Understanding humorous episodes often requires the comprehender to draw a number of inferences about the characters and events in the episode. Unlike comprehension of other types of material, however, these inferences often are not straightforward nor applied "automatically." Instead, a logical series of inferences may be required, implying a form of problem solving activity. This characterization of comprehension of humorous material is particularly appropriate for multiple frame cartoons. Since we are especially interested in the role of these sorts of cognitive processes in understanding humorous material, we have focused our efforts on cartoon comprehension.

The process model that we have been trying to develop is based on recent conceptualizations of comprehension processes and on theoretical ideas about the defining attributes of humorous material. We characterize a cartoon as a brief narrative that is described pictorially and linguistically. As such, much of the information learned from reading a cartoon is inferred rather than explicitly stated or illustrated. It is further assumed that as a cartoon is processed expectations about upcoming information are constructed, just as in reading ordinary discourse. Often these expectations are violated by some form of incongruous information, causing a revision of previously inferred information or at least a reorganization of knowledge about the event being described. This revision or resolution of incongruity has been claimed to be an important element of humor by recent theorists. In addition, resolution may introduce further incongruities which are never resolved. We argue that these inferential problem solving processes are of central importance in any account of comprehension or appreciation of humorous material.

In order to specify in more detail the nature of comprehension processes associated with reading cartoon material, we made use of multiple regression modelling procedures. We planned to focus on cognitive processes involved in cartoon comprehension and, to a lesser extent, on the characteristics of cartoons that are associated with humor. Therefore, we sought to develop regression models that would account for variability of cartoon items on three dimensions: (a) comprehension time, (b) problem solving processes associated with understanding the joke on which a cartoon is based, and (c) degree of humor.

We selected a set of 64 cartoons sampled from five different syndicated cartoon strips. The items were originally published between 1976 and 1978. A sample of 35 university students rated (on a 7-point scale) the degree of difficulty involved in understanding or figuring out the joke on which each cartoon was based. This served as a very general measure of the complexity of the problem solving comprehension processes we have postulated. Another sample of 56 students was shown the same items in the following manner. First, all frames of a cartoon except the last frame were shown together, then the last frame was shown. Time required to comprehend each section (first frames and last frames) was measured. In addition, these subjects rated on a 7-point scale the degree of humor of each cartoon.

Cartoons were then classified according to a set of independent variables representing many characteristics of cartoon items. Verbal information variables included number of letters, syllables, words, propositions, and sentences and number of speaker transitions. Visual information was coded by such variables as number of characters, number of physically transformed characters, change of scenes, activity, and presence of a visual detail which was critical for understanding a cartoon. Cartoons were also coded according to whether a possible or impossible incongruity was introduced in the last frame and left unresolved. The logic on which a cartoon was based was also coded. Categories included exaggeration, iteration (e.g., a tow truck towing a tow truck), and model theory (the reader interprets an ambiguous statement or situation in the most usual way, then finds that an alternative interpretation was intended). Finally, in order to reflect cartoonists' style differences, the items were classified according to cartoon strips.

Four regression analyses will be reported. In all cases a stepwise multiple regression was used in which variables were added into the regression equation until an optimal adjusted R^2 was reached. The first two analyses involved mean comprehension time on first frames and on final frames. The list of independent variables included those mentioned above as well as mean rated humor and difficulty in understanding the cartoon's joke. We had two general expectations about these analyses. First we expected to do a better job of accounting for variability in first frame comprehension times since most of the complex comprehension processes are associated with the last frame. For similar reasons we expected the last frame comprehension time regression equation to more heavily weight the difficulty rating variable than would the equation for first frames.

The optimal adjusted R^2 for the first frame analysis was a very respectable .94 (multiple R=.98). Eighteen variables were entered although six did not contribute significant predictive power. Comprehension time increased as a function of number of letters, sentences, and transformed characters in the first frames. In addition, relatively small contributions were made by difficulty rating (smallest beta weight of all significant variables) and the variable which coded exaggeration: longer comprehension times were associated with higher difficulty and cartoons based on exaggeration. The seven other significant variables were all supressor variables: very low simple correlations with comprehension time, but they accounted for some variability of significant predictor variables which was not associated with comprehension time. Most of these suppressor variables were associated with verbal and visual information contained in the last frame.

The regression analysis of first frame comprehension time successfully accounted for a great majority of variability. Material of this sort does not normally call for much complex cognitive activity while reading the first frames, and so a few simple variables are sufficient to account for most of the variance. Number of letters and transformed characters together produced a regression equation with an adjusted R^2 of .86. Rated difficulty did not play a large role in accounting for first frame comprehension time. Finally, it was rather surprising that other visual information variables such as number of characters did not have much impact. It appears that as long as characters maintain a consistent appearance readers require very little time or effort to recognize them, even when first introduced.

The regression analysis of final frame comprehension time produced an optimal adjusted R^2 of .79 (multiple R .91), with 13 independent variables involved in the equation. Only five variables made significant contributions. Comprehension time increased with higher numbers of letters in the last frame, rated difficulty, number of new characters introduced in the final frame, and number of syllables in the first frames. Number of syllables

in the last frame was a supressor variable. As expected, the importance of problem solving comprehension processes was evident in the analysis. Rated difficulty was the second variable entered into the equation for last frame comprehension time. In addition, the introduction of new characters in the last frame requried more reading time. Reinterpreting information from the first frames, or using that information to understand the last frame are important processes since amount of first frame verbal information strongly influenced last frame comprehension time.

Since difficulty ratings played such an important role in the comprehension time analyses, it would be interesting to discover which variables contribute to the degree of difficulty in comprehending a cartoon. The relevant regression analysis indicated that the present data could offer only limited aid in answering this question. The optional adjusted R^2 was only .52 (multiple R = .85), with 26 variables in the equation. Thirteen of the variables made significant contributions. Cartoonists' style played an important role as the first variable entered was the "0/1" variable which coded cartoons as being from "Wizard of Id" versus some other cartoon strip. According to the regression model, such cartoons added a full 1.5 units of difficulty on the 7-point scale. Cartoons based on iterative logic also increased difficulty. Increases in the number of propositions, syllables, or speaker transitions in the first frames also increased difficulty ratings. The other seven variables (other cartoons and last frame information) acted as suppressor variables. This analysis clearly points to the importance of reviewing first frame information while trying to comprehend the final frame and basic joke of a cartoon. Difficulty was not associated with amount of final frame information which, in itself, is just as straightforward as verbal and visual information contained in the first frames. Interpreting the last frame in the context of previous information, however, is a major part of the reader's problem solving comprehension process.

Finally, what makes cartoons funny? The regression analysis of humor ratings produced a rather low adjusted R^2 of .60 (multiple R = .82), so no definite answers are yet available. Nine significant variables were involved, two of which are related to cartoonists' style. The "Wizard of Id" and "Hagar the Horrible" were associated with rather high humor ratings. These ratings also increased with the number of scene changes in the first frames and the number of speaker transitions in the last frame. Adding an impossible incongruity in the last frame seemed to improve the humor of cartoons, but basing a cartoon on a critical visual detail in the final frame was associated with reduced humor ratings. The other three variables were suppressor variables. Obviously this analysis does not provide a very satisfactory explanation of what makes a cartoon humorous, but it has helped to identify attributes that characterize successful attempts to create humorous materials.

Our future efforts will be directed at a closer and more detailed study of the problem solving comprehension processes involved in reading cartoons. In addition, it is clear that it will be a difficult task to identify the nature of the relationship between humor and these cognitive processes. In our data there is presently no clear relation (linear, or otherwise) between humor and difficulty ratings. Although we have begun to make progress in the development of a model of cartoon comprehension processes, the secret of the humor of cartoons is still well kept.

Demon Timeouts

*Limiting the Life Span of Spontaneous Computations in Cognitive Models**

Steven Small

Department of Computer Science
University of Rochester
Rochester, New York 14627

&

Departement d'Informatique
Universite Paris VIII Vincennes
93200 Saint Denis FRANCE

Introduction

The limitations in human short-term memory, of whatever extent they are determined to be, must be accounted for in constructing cognitive models. The view of this memory as a data (or knowledge) store leads to models with a fixed-length buffer of quickly accessible *short-term* data, and general procedures for replacing items in that buffer with other items (see Marcus [1979], for example). Another vantage point on this question comes from viewing the this memory in terms of the active problem solving processes of the model. This perspective suggests building models with time-limited processes that may not remain active indefinitely. Just as the size of an individual constituent in a fixed-length buffer may depend on the nature of that item, so the time limit of an active process (even the units of time employed) must be allowed to depend on the nature of its functioning.

Every active process in a cognitive model must be associated with another process, called a *timeout process*, which can force its premature termination (its *timeout* or *expiration*) on specific criteria. The timeout specifies how long the process (the *parent process*) may remain active, in terms of specific time units or model events, and in addition, what actions to take (or processes to initiate) on expiration. Note the enormous significance of this organization: The processing of individual actors in a model can depend on either the existence or the non-existence of relevant information. In terms of human problem-solving, this is analogous to our basing some decision on the explicit knowledge that we do not know something. Often the knowledge that we do not know something can be as useful to solving a problem as any other sort of knowledge.

*The initial research motivating the ideas described in this report was supported by grant NSG-7253 from the National Aeronautics and Space Administration to the University of Maryland. During the writing of this paper, the author was supported by a Fulbright-Hayes Lectureship in Artificial Intelligence at the Universite de Paris, and computer facilities were provided by the *Institut de Recherche et Coordination Acoustique/Musique*. The support of NASA, the Fulbright-Hayes program, and IRCAM are gratefully acknowledged.

Word Expert Parsing

This kind of approach has been taken in the modelling of human language comprehension by the *Word Expert Parser* (WEP). WEP views language comprehension as interactions among a large number of human cognitive processes, including word-based context analysis processes called *word experts*, and other mechanisms such as belief maintenance, general problem solving, and so forth. The system models this *understanding as memory interactions* through the message-passing behavior of its different constituent processes. Evaluation of the system ought not to be made through its output; rather, its important aspects are (a) the organization and representation of the model processes, (b) the order of execution of dependent processes and the concurrence of others, and (c) the messages exchanged among the processes.

An important aspect of WEP is that its processes are not permitted to wait indefinitely for information they might need to perform their functions. In understanding the meaning of a word in some sentence, for example, often a reader requires some knowledge of the context following it. How long would a reader wait for a piece of disambiguating information before choosing one of the competing interpretations of the word? Does this wait depend on general syntactic or semantic criteria, or on the particular words involved and their idiosyncratic couplings with other words? And a last question: How can a model of language understanding (or of any cognitive mechanism) account for this limitation in active short-term memory processing?

As a completely procedural distributed model, WEP does not contain some finite fixed length buffer to represent a short-term memory, but models the memory limitation with processes that have a strictly limited life span. The human understander does not wait indefinitely for information that could aid his/her understanding, and the model cannot either. When a process in WEP decides to await some piece of information from another process, a pattern-invoked *timeout process* is initiated to monitor the duration of the wait. In the ideal case, the desired information quickly becomes available in the model, and the awaiting process promptly receives it. If the information does not appear within a certain time, however, the timeout process must command the awaiting process to go on without it.

Timeouts

The general notion behind these *timeouts* is a simple one. Anytime a cognitive model would ordinarily create a demon, or any pattern-invoked process (a *spontaneous computation* [Rieger, 1977]), it should instead create a pair of competing demons, a *parent* demon and a *timeout* demon. The timeout demon should be guaranteed to trigger within a certain fixed amount of processing by the model as a whole. If the parent triggers first, the timeout disappears and the parent process performs its pattern-invoked operations. If the timeout triggers first, there are two possibilities; either (a) both the timeout and the parent disappear immediately, or (b) the timeout disappears immediately and the parent performs certain special timeout operations and then disappears. The timeout operations are process-specific and only execute in the case when their associated pattern-invoked process (demon) does not trigger within its allotted life span.

The result of this arrangement is to place a limitation on the duration of applicability of demons in the model, preventing them from waiting indefinitely for desired information. In this way, futile waits for information that will never arrive can be avoided, and excessively long waits -- far exceeding human memory limitations -- can be disallowed. Furthermore, by appropriate use of timeout operations, processes in a cognitive model can be triggered by either the invocation or the non-invocation of some pattern in the memory. By not knowing something within a certain time limit of model events, and by knowing about this lack, a system can make valuable inferences.

Timeout Criteria

The question of how long to spend looking for helpful information before going on without it, whether in problem-solving or cognitive modelling, is clearly not a new one. How far ought one to search a promising path in a chess tree, or how many associations in a semantic memory ought to be examined, before available knowledge is used to come to some conclusion, even if not the right one? Various answers to this problem have been studied, although from a different perspective than the one suggested here. From our vantage point, there are two interrelated questions: (a) What objects or events might be used to trigger demon timeout in a particular cognitive model? And similarly, (b) What sequence of events in the model can be used to measure the duration of the wait for desired information?

The progression of events in a cognitive model depends fundamentally on the nature of the particular activity being modelled. In natural language understanding, a person reads words and sequences of words of various lengths, creates conceptual conglomerates, follows character and plot development (in reading a story) or speaker intentions (in participating in a dialogue), makes hypotheses and has some hypotheses fulfilled and others rejected, and so forth. Each of these facets of language understanding represents an event that might be used to trigger demon timeout in an understanding model. Furthermore, the events could be counted and treated as units of time for measuring durations in the model.

For a model of government decision-making or administrative behavior, perhaps the number of memoranda reaching a particular department would be a good gauge on the progress of the problem-solving activity. Another measure might be the number of responses from departments participating in a decision, or from sub-departments of a modelled one. Recent history suggests that in modelling government processing, the number of adverse newspaper articles or adverse decisions in other branches of government could be useful. Problem-solving systems must not wait indefinitely for data needed to make inferences; better make a wrong inference than come to no decision at all. The impatience of people in such decision-making situations makes this all the more important in cognitive modelling.

Timeouts in Word Expert Parsing

The processes in the Word Expert Parser limit their wait for relevant information by employing several of the yardsticks mentioned above; i.e., the system maintains counts on certain model actions intended to correspond to relevant events of natural language understanding. At present, these include the total number of words read by the system, the number of conceptual conglomerates created by the model, and the number of sentence breaks encountered. While these are fairly straightforward measures, our current research will give WEP two new timeout criteria. In both cases, the timeout processes might be considered meta-processes, as their actions are based on the functioning of other processes (rather than on their results or on messages from them).

The first new waiting criterion causes a timeout on the termination of some model process. Thus, a process might wait for some information just as long as some other process were still active (and could provide it, for example). The moment the specified process terminated, the demon for the awaiting process would expire, and perhaps carry out its special timeout actions before terminating itself. This timeout criterion has been simple to implement. The other new waiting criterion causes the timeout of some demon on the creation of an identical demon (i.e., one awaiting the same information) by another process. To put that another way, the system may permit a process to wait for some datum or event just so long as no other process initiates a wait for the same datum or event.

This particular strategy aids in modelling a phenomenon that occurs frequently in natural language comprehension. The problem of anaphoric reference illustrates the issue. When a computer model such as WEP begins reading a definite noun group, it ought immediately to initiate a reference process to search for the referent of that group. How long should this reference process be permitted to search before using available information to conclude its processing? One answer might be that when the understanding of further discourse required knowledge of the referent, the process should come to some decisions immediately. The timeout actions of a reference process might provide some default substitute for the desired referent, for example, while at the same time initiating some new process to proceed along a different path based on the (known) situation at hand.

The timeout of a process allows it to come to some conclusion without waiting indefinitely for data that will insure the correct decision. In this example, the timeout of the reference process takes place when the understanding of subsequent text requires information about the searched-for item. In this way, the reference process can be forced to make some inference (e.g., the definite noun group has no referent) when the overall understanding process requires it, but not before. Ideally, of course, the correct referent would be found before then, and none of this would be necessary. Timeout demons exist, however, solely for those situations when the ideal context for making correct inferences does not present itself. They are a means for deciding when to make an inference without all the facts being present.

Summary

An important facet of general problem-solving, and likewise cognitive modelling, is how long to spend looking for helpful information before going on without it. In a computer model based on spontaneous pattern-invoked processes (demons), there must be a means for limiting the duration (in terms of some measure of model convergence) of their applicability. This paper suggests *timeout processes* as a means of limiting the life span of pattern-invoked demons in a cognitive model. Every process in a model must be associated with a timeout process, which monitors various specific model events and prevents the process from remaining active indefinitely. If the process does not finish its activity before being stopped by its timeout, it may nonetheless perform certain timeout actions before disappearing. These actions can make use of the (possibly valuable) information that the trigger pattern for the process never arrived. Timeouts thus enable processes to use knowledge about not having certain knowledge to help make important inferences in cognitive models.

References

Marcus, Mitchell P. (1979), *An Overview of a Theory of Syntactic Recognition for Natural Language*, AI Memo #531, MIT Artificial Inteligence Laboratory.

Rieger, Chuck (1977), *Spontaneous Computation in Cognitive Models*, Cognitive Science, vol. 1, no. 3.

Small, Steven (1980), *Word Expert Parsing: A Theory of Distributed Word-Based Natural Language Understanding*, TR #954, Department of Computer Science, University of Maryland.

Small, Steven (1981), *Toward a Cognitive Mechanics for Distributed Modelling*, Technical Report, Department of Computer Science, University of Rochester (to appear).

The effects of integrated knowledge on fact retrieval and consistency judgments: When does it help, and when does it hurt?

Lynne M. Reder
Carnegie-Mellon University

and

Brian H. Ross
Stanford University

To understand how we can easily retrieve facts from memory we must also understand the limitations of memory. Interference has been recognized for a long time as a major source of forgetting. More recently interference has also been shown to affect speed of retrieval. People are slower to recognize a studied fact when other studied facts share some of the same concepts (e.g., Anderson, 1974, 1976; Anderson & Bower, 1973; Hayes-Roth, 1977; King & Anderson, 1976; Lewis & Anderson, 1976; Thorndyke & Bower, 1974). This interference effect has been called the fan effect by Anderson (1974), because of the underlying representation he assumes to explain this interference phenomenon. Specifically, facts are assumed to be represented in a network structure where nodes in the network represent concepts, and links represent relations among concepts. Studied facts that share the same concepts would be represented as propositions with relational links fanning out of the same concept nodes. Time to retrieve any one fact depends on the time to activate the entire proposition. Activation spreads from various concept nodes until the entire proposition is activated, and it is slowed down when there are more links that divide the activation (see Anderson, 1976, or Collins & Loftus, 1975).

This robust fan effect has been under close scrutiny of late (e.g., McClosky & Bigler, 1980; Moeser, 1977, 1979; Smith, Adams, & Schorr, 1978) due to the paradoxical implications of the effect. The paradox is as follows: An expert knows more about the topic than a novice, and we would expect an expert to be able to answer questions about that topic faster than a novice. Yet according to the theory, the more one knows about a particular topic, the more potential interference to that topic, and the slower should be retrieval of any given fact.

Smith, Adams, and Schorr (1978) and Moeser (1977; 1979) partially demystified this paradox by showing that when the facts associated with the particular concept are themselves thematically related, there does not seem to be any interference among them with respect to time to verify the truth of one of these facts. Reder and Anderson (1980) showed that this attenuation of the fan effect only occurs when subjects can make plausibility judgments rather than explicit fact retrieval judgments (e.g., when foils are not thematically related to the facts studied). Reder and Anderson propose a propositional network with the additional assumption that related facts are stored in a subnode structure attached to individual nodes. Recognition and consistency judgments use the same representation and activation process, but require the subjects to use different criteria. When subjects are not forced to retrieve a specific fact, because the foils are not thematically related, subjects stop search at the appropriate thematic subnode. This subnode model accounts for another Reder and Anderson finding: Subjects were slower to verify a fact the more themes there were associated with the fictitious individual. The interference effects did not depend on the number of facts about the themes unrelated to the one specifically queried, only the number of irrelevant themes. This model also explains a similar result of McClosky and Bigler (1980). There is other empirical support for the notion that people will answer a question by judging plausibility, when asked to make recognition judgments, (Reder, 1980, 1981). In everyday situations, people make plausibility judgments rather than trying to decide if a specific fact has been presented to them. We usually do not have to discriminate something that was said to us from something that is true but is an inference or a paraphrase of something that we heard.

These earlier studies suggest the following tentative conclusions. Regardless of the nature of the foils, the fan effect is attenuated by the subnode structure afforded by thematically related material. That is, having subnodes speeds search by pruning search at inappropriate subnodes. Therefore, fan is not computed for the total number of facts learned about a concept. Moreover, when the foils are not thematically related to the learned facts, subjects can also stop search at the appropriate subnode, and circumvent the effect of "relevant fan" (the number of facts associated with the probed theme). However, regardless of whether subjects use a plausibility judgment or must retrieve a specific fact to answer a question, there is still a fan effect for the number of themes associated with a concept, i.e., the number of links from the concept node to the various subnodes.

There were two basic motivations for the research to be reported in this paper: The first is to extend and replicate previous research to insure that there really is a fan effect for number of themes. The second is to examine the effect of relevant fan when subjects are asked to make consistency judgments. We expect a negative fan effect in the consistency condition rather than an attenuation of the effect. The more relevant facts that are available from which to select in order to make a consistency judgment, the faster a subject can find enough information to make a judgment. The fact that there was only an attenuated fan effect with unrelated foils may reflect a mixture of plausibility judgments and direct retrievals when asked to make a recognition judgment.

In addition to the recognition blocks, (one block with thematically related foils and one with unrelated foils), and the consistency judgment blocks, we also included another kind of judgment which we call theme judgments. On these trials subjects see the fictitious individual's name (e.g., Marty) and a theme or topic name (e.g., ship-christening). We expect subjects to stop at the theme node for both types of judgments. The latter condition serves as a check on the processes used for consistency judgments.

Method

Procedural Overview[1]

There were two major phases in the experiment: study and test. In the initial study phase, subjects learned sets of facts about various characters. This phase included an initial presentation of facts for each character, organized by theme. Then subjects studied and were tested on the materials using two sets of dropout procedures, so that they could recall the facts studied about each character. Subjects with below a 90% criterion of final recall were excluded from the analyses.

In the critical phase, the test phase, reaction times to make various types of judgments about the learned material were collected, with each judgment type tested in a different block. The three judgments were recognition judgments with foils thematically related to the sentences actually studied about the probed character, recognition judgments with foils unrelated to the material studied about the probe character, and consistency judgments (whether the probe was consistent with what had been studied). Intermixed among these various test blocks were theme judgments, where subjects would see a character name and a theme name, rather than an entire sentence.

Both speed and accuracy were emphasized. Subjects were told to respond as fast as they could while remaining very accurate. Feedback was given after every trial.

[1] This description must be brief due to space limitations, for a fuller description, consult Reder and Ross (in preparation).

Table 1

Examples of Studied Facts and Test Questions

# of facts in the three themes	
3-2-1	Alan bought a ticket for the train. Alan heard the conductor call "All aboard". Alan arrived on time at Grand Central Station
	Alan added bleach to the rinse cycle. Alan sorted his clothes into colors and whites.
	Alan fell while skiing down the steepest stretch.
1-1-0	Brian watched the freaks in the side show.
	Brian wanted to major in psychology.
4-0-0	Steven called to have a phone installed. Steven read and signed the lease. Steven unpacked all of his boxes. Steven mailed out change of address cards.
3-3-0	James compared 5 different model cars. James paid the car dealer in cash. James put the license plates on his car.
	James checked the Amtrack schedule. James arrived on time at Grand Central Station. James watched the trains from the platform.

Recognition	Consistency	
Target	Target (Yes)	(3-2-1) Alan bought a ticket for the train.
Foil Unrelated	Inconsistent (No)	(1-1-0) Brian unpacked all of his boxes.
Foil Related	Thematic (Yes)	(3-3-0) James bought a ticket for the train.
Theme judgment (True)		(3-2-1) Alan train
Theme judgment (False)		(3-2-1) Alan circus

Design and Materials

Table 1 illustrates some of the material that a subject might see. There were several factors that define the condition within which a particular probe is tested: relevant fan (one to four facts related to the probe), theme fan (one to three themes learned about each character), and irrelevant fan (zero to five facts, irrelevant to the probe, that were also learned about the probed character).

The test variables were type of judgment required (recognition with related foils, with unrelated foils, consistency judgments, and theme judgments) and probe type (target, foils, and thematic, the latter defined only for consistency judgments). Some examples are illustrated in Table 1. Study and foil materials were constructed for each of the 33 subjects randomly.

Results

Reaction times were truncated to 5 seconds and missing cells replaced with RTs of 5 seconds. An analysis of variance was performed separately on each of the eleven task types because the variances are not the same in the different tasks. The data have been collapsed in different ways to analyze different aspects of the experiment, each of which can only be summarized here.

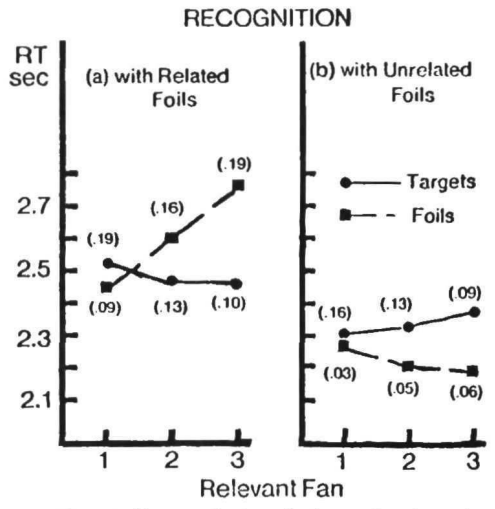


Figure 1. Mean reaction times (and proportion of errors) as a function of relevant fan in the recognition blocks.

First consider the recognition blocks, shown in Figures 1 and 2 (with the percent error listed above each point). There is no effect of relevant fan in the unrelated foil block. For the thematic foils, reaction time increased with relevant fan and, as in Reder and Anderson (1980), the targets in this block showed an effect of relevant fan on accuracy. As the number of themes associated with the probed character increased, reaction time increased for the targets in both recognition blocks, but not for the foils.

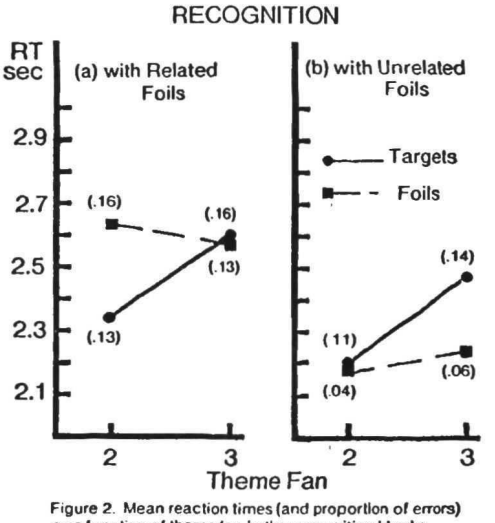


Figure 2. Mean reaction times (and proportion of errors) as a function of theme fan in the recognition blocks.

Now, consider the results from the consistency blocks of trials, shown in Figures 3a and 4a. Reaction times are plotted only for target (presented) sentences and for thematically related sentences that were not studied, because the unrelated statements cannot be plotted as a function of relevant fan. There is a significant negative fan effect for thematic statements, such that subjects are faster to make a consistency judgment the more facts they know on the relevant topic. Reaction times also decrease for targets in the consistency block. There is a significant effect for both targets and themes on accuracy, such that subjects are also more accurate the greater the relevant fan. The number of themes associated with the probed characters, collapsed over the three probe types, show a fan effect.

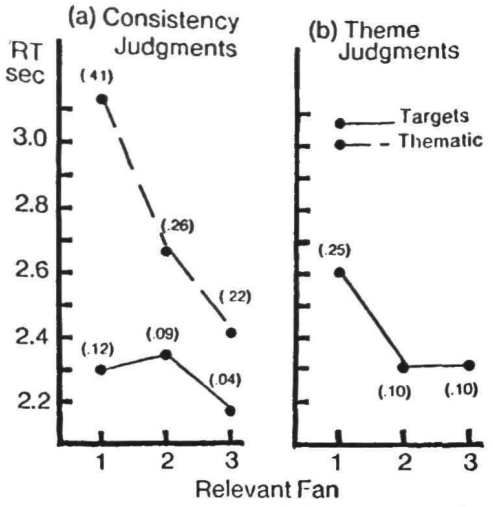


Figure 3. Mean reaction times (and proportion of errors) as a function of relevant fan in the consistency block and for the theme judgments.

Theme judgments (Figures 3b and 4b), like consistency judgments show a significant negative fan effect for the relevant theme for the positive judgments. There is a strong positive fan effect of number of themes. In no task was there an effect of the irrelevant fan on reaction time.

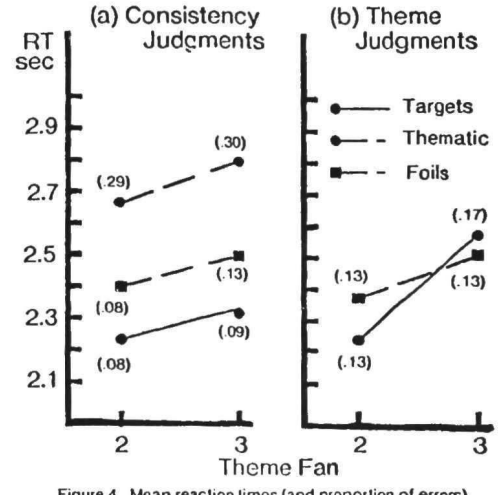


Figure 4. Mean reaction times (and proportion of errors) as a function of theme fan in the consistency block and for theme judgments.

Discussion

The conclusions that can be drawn from this pattern of data are straightforward. First, we have a better understanding of the conditions under which the fan effect is obtained. The fan effect is obtained only when subjects cannot use plausibility to answer the question, when subjects are asked to find a specific fact in memory and the foils are thematically related. In other situations, the strategy is to try to decide if a statement is true using consistency judgments, regardless of whether or not subjects are asked to make consistency judgments or in fact asked to make recognition judgments. This result has been found elsewhere (Reder, submitted; Reder & Anderson, 1980).

The suggestion of a fan effect due to the number of themes (McCloskey & Bigler, 1980; Reder & Anderson, 1980) has been confirmed and extended. We found, when controlling for total number of facts studied, that the number of themes associated with the character positively affects reaction time.

More important, perhaps, is the finding that fan facilitates question-answering when exact fact retrieval or recognition is not needed, in the consistency and theme judgments. This is a better resolution of the paradox of the expert than a finding that fan need not hurt retrieval time. Rather, we have found that knowing more actually speeds decision time in many situations. Our explanation for a negative fan effect (see Reder, submitted, and Reder & Ross, in preparation, for a more thorough explanation)is that the more facts attached to the subnode the stronger the link to that subnode. In other words, rate of activation is not only a function of the number of links to a concept but the relative strength of various links. Strength of a link is a function of its usage (usage depends on the frequency that the link is traversed), and recency of last usage. In fact, the experiments reported here manipulated strength independently, and found support for this notion of differential strength of arcs affecting response time. For space considerations, that result was not discussed here. Those results will be described in detail in Reder and Ross (in preparation).

In summary, we have learned that experts are not hampered by knowing too much for several reasons: first, experts organize their information into more specific subtopics. This is reasonable, because experts understand their topic area well enough to appreciate subcategories. Second, experts are not asked whether they recognize having been told a specific fact; rather, they are asked to judge whether something is true or something is plausible. Therefore, they can use their redundant knowledge in order to speed judgment.

References

Anderson, J.R. Retrieval of propositional information from long-term memory. Cognitive Psychology, 1974, 5, 451-474.

Anderson, J.R. Language, Memory, and Thought. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Anderson, J.R., and Bower, G.H. Human Associative Memory. Washington: Winston & Sons, 1973.

Collins, A.M., and Loftus, E.F. A spreading-activation theory of semantic processing. Psychological Review, 1975, 82, 407-428.

Hayes-Roth, B. Evolution of cognitive structures and processes. Psychological Review, 1977, 84, 260-278.

King, D.R.W., and Anderson, J.R. Long-term memory search: An intersecting activation process. Journal of Verbal Learning and Verbal Behavior, 1976, 15, 587-606.

Lewis, C.H., and Anderson, J.R. Interference with real world knowledge. Cognitive Psychology, 1976, 7, 311-335.

McCloskey, M., and Bigler, K. Focused memory search in fact retrieval. Memory & Cognition, 1980, 8(3), 253-264.

Moeser, S.D. The role of experimental design in investigations of the fan effect. Journal of Experimental Psychology: Human Learning and Memory, 1979, 5, 125-134.

Reder, L.M. Plausibility Judgments vs. Fact Retrieval: Efficient Strategies for Question-Answering. Paper presented at meetings of the Psychonomics Society, Nov. 1980.

Reder, L.M., and Anderson, J.R. A partial resolution of the paradox of interference: The role of integrating knowledge. Cognitive Psychology, 1980, 12, 447-472.

Reder, L.M., and Ross, B. A resolution of the paradox of interference: The role of integrated knowledge in fact retrieval and consistency judgments. (In preparation.)

Smith, E.E., Adams, N., and Schorr, D. Fact retrieval and the paradox of interference. Cognitive Psychology, 1978, 10, 438-464.

Thorndyke, P.W., and Bower, G.H. Storage and retrieval processes in sentence memory. Cognitive Psychology, 1974, 5, 515-543.

A QUESTION ANSWERING METHOD OF EXPLORING PROSE COMPREHENSION: AN OVERVIEW

Arthur C. Graesser
California State University, Fullerton

When individuals comprehend prose, they construct a large number of inferences and expectations. Where do these inferences and expectations come from? It is believed that generic schemas provide the background knowledge that is needed to generate inferences and expectations. The schemas correspond to different knowledge domains and levels of structure. There are schemas for objects, actors, event sequences, goal oriented activities, and so on. During comprehension the comprehender identifies a variety of schemas. When a schema is identified, it guides the interpretation of explicit input as well as the construction of inferences and expectations.

There are several schema-based models in psychology and other disciplines in cognitive science. Psychologists have typically investigated global issues regarding schemas in comprehension. They have rarely specified detailed representations and symbolic procedures. However, during the last four years the Cognitive Research Group at Cal State Fullerton has ventured into a detailed and very time-consuming project. We set out to explore the following problems:

(1) To identify the inferences and expectations that comprehenders generate when a passage is comprehended.

(2) To trace the constructive history of specific inferences and expectations when a passage is comprehended on-line.

(3) To formulate a system for representing knowledge.

(4) To map out the content and structure of schemas which are invoked when a passage is comprehended.

(5) To examine how conceptualizations in generic schemas are passed to the representation of a specific passage.

(6) To assess whether behavioral data can be explained by properties of passage representations and the process of constructing these representations. The behavioral tasks include question answering, recall, and inference verification.

We have analyzed both narrative and expository passages. However, the most extensive analyses have been on short narrative passages such as the following:

The Czar and His Daughters

Once there was a Czar who had three lovely daughters. One day the three daughters went walking in the woods. They were enjoying themselves so much that they forgot the time and stayed too long. A dragon kidnapped the three daughters. As they were being dragged off, they cried for help. Three heroes heard the cries and set off to rescue the daughters. The heroes came and fought the dragon and rescued the maidens. Then the heroes returned the daughters to their palace. When the Czar heard of the rescue, he rewarded the heroes.

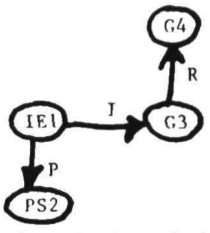
Conceptual Graph Structures

It is believed that knowledge can be represented in the form of conceptual graph structures. A graph structure is constructed when an individual comprehends a passage such as the Czar story. Both explicit statements and inferences are structurally interrelated in a graph structure. Similarly, the content of generic schemas are represented as conceptual graph structures.

In our representational system, a graph structure is a set of labeled statement nodes which are interrelated by labeled, directed arcs. A statement node is roughly a proposition. Each statement node is assigned to one of six node categories: Physical State, Physical Event, Internal State, Internal Event, Goal (which includes actions), and Style. For example, three heroes heard the cries and set off to rescue the daughters has the following four statement nodes:

(1)Internal Event: heroes heard cries
(2)Physical State: there were three heroes
(3)Goal: heroes set off
(4)Goal: heroes rescue daughters

There are five categories of labeled, directed arcs: Reason(R), Initiate(I), Manner(M), Consequence(C), and Property(P). These arcs interrelate the nodes. The above four statements would be interrelated as follows:



For a more complete description of this representational system, see Graesser (1981) and Graesser, Robertson, and Anderson (1981).

One distinctive feature of this representation is that it captures differences between goal-oriented structures (with Goal nodes, Reason arcs, and Initiate arcs) and causally-oriented structures (with Event nodes, State nodes, and Consequence arcs).

A Question Answering Method of Exposing Implicit Nodes

A question answering (Q/A) method has been used to expose inferences and expectations. After comprehenders read a passage, they answer why, how, and what-happened-next (WHN) questions about each explicit statement in the text. The answers include a large number of implicit nodes. An implicit node is analyzed if it is produced as an answer by at least two comprehenders. Thus, the ideosyncratic answers are eliminated.

When comprehenders are probed with questions after reading a passage, the inferences in the Q/A protocols are classified as preserved nodes. There is a large number of preserved nodes in a short passage such as the Czar story. In our analyses, there were 189 preserved nodes. Therefore, there were 10 unique inference nodes for every explicit node. Conceptual graph structures were composed from the set of explicit nodes and inference nodes.

Constructing Graph Structures On-line

It is possible to trace the constructive history of each node in a passage structure. This is accomplished by manipulating the amount of passage context that a comprehender receives before a statement is probed with questions. In a No Context condition, passage statements are probed out of context. The inferences generated in this condition are classified as statement-driven (SD). In a Prior Context condition, passage statements are probed on-line; the comprehender reads only the passage content up through the target statement and then the statement is probed. Inferences in this condition are classified as prior-context-driven (PCD) if they are not SD. In a Full Context condition, the comprehender reads the entire passage before the passage statements are probed. Inferences in this condition are subsequent-context-driven (SCD) if they are neither SD nor PCD.

On the basis of these context manipulations we traced the constructive history of each inference node (answer to why or how question) and expectation node (answer to WHN question). Each implicit node was classified on the basis of the first explicit node in the passage which activated the implicit node. Of the preserved inference nodes, 61% were SD, 34% were PCD, and only 5% were SCD. Consequently, the on-line representation of a target statement accounts for 95% of the inferences associated with the statement; subsequent context adds very little.

Sometimes implicit nodes are generated sometime during comprehension , but are later disconfirmed. There were 108 disconfirmed nodes in the Czar story. Erroneous SD nodes were usually disconfirmed by subsequent context, rather than being blocked by prior context.

We traced the entire evolution of the conceptual graph structures as passage statements are interpreted incrementally, statement by statement. This analysis included both preserved nodes and inference nodes. The analysis of the Czar story revealed that established conceptualizations are rarely restructured as new information is received. New nodes were rarely inserted inbetween old nodes; erroneous old nodes were rarely removed from chains of old nodes that end up being preserved. Instead, new nodes were appended to old nodes; erroneous nodes and node chains were pruned from old structures. A pruning+appending mechanism explained much of the on-line construction of graph structures for narrative.

Schema Content and Structure

A free generation plus Q/A method has been used to map out the content and structure of schemas. Consider a DAUGHTER schema. Individuals in a free generation group write down typical actions and attributes of daughters. Free generation nodes include all statements that are produced by two or more individuals. A second group of individuals participate in a Q/A task. Each of the free generation nodes are probed with a why and a how question. The final set of nodes include all statements produced by at least two individuals. Conceptual graph structures are then prepared for schemas such as the DAUGHTER schema.

We have identified 31 schemas that are relevant to the Czar story. The content of these schemas has been analyzed using the free generation plus Q/A method. Twenty of these schemas were classified as microstructure schemas because they referred to explicitly mentioned actors, objects, actions, or properties (e.g., HERO, PALACE, KIDNAP, ATTRACTIVENESS). There were 11 macrostructure schemas which were derived from the text on the basis of our intuitions (e.g., GOODNESS, FAIRYTALE, RETURNING FAVOR). The number of nodes per schema varied from 32 to 187 with a mean of 93 nodes. The number of free generation nodes varied from 3 to 22 with a mean of 13. Therefore, the Q/A task exposed most of the schema content.

Passing Schema Nodes to Passage Representations

Among the 31 schemas relevant to the Czar story, there were 2883 nodes. Only 362 of these generic nodes matched an inference node in the passage. There were three types of matches between inference nodes and schema nodes. Exact matches accounted for 9% of the matches. Most of the matches (86%) involved an argument substitution, as shown below:

Schema node: person get exercise
Inference node: daughters get exercise

Some matches (5%) were more complex. Since only 13% of the schema nodes were passed to the conceptual graph structure for the passage, a substantial number of schema nodes are somehow eliminated. We are presently examining symbolic mechanisms that might explain which generic nodes are passed to the passage structure.

The schemas accounted for most of the nodes in the Czar story structure. For 79% of the implicit passage nodes, there was a match with a node in at least one schema. For 74% of the inference nodes, there was a match with a node in at least one microstructure schema; there was a match with a node in at least one macrostructure schema for 31% of the inference nodes. On the average, an inference node had a match with 1.7 schemas. We are presently examining how schema structures are synchronized with structures of passage excerpts--as passage statements are comprehended on-line.

Question Answering, Recall, and Verification Ratings

Do the conceptual graph structures correspond closely to human conceptualizations? One way to assess this is to examine whether the structures explain behavioral data. We have extensively examined patterns of data in three tasks: question answering, recall of explicit information, and verification of nodes in the graph structures (see Graesser, 1981; Graesser et al., 1981). The results of these analyses have been encouraging.

We assumed that specific symbolic procedures are invoked in any given behavioral task. Most of the symbolic procedures were written in the form of a production system. A production system operates on a conceptual graph structure and thereby generates expected output. Hopefully, the expected output would match closely to the obtained behavioral output. For example, we formulated production systems for specific types of questions. The production system for one type of question follows different paths of arcs and nodes than does that of another type of question; the expected answers would therefore be different for the two types of questions. The production systems and graph structures have together accounted for 90% of the specific answers to specific questions.

Final Comments

We have been impressed with the Q/A method as an empirical technique for exploring comprehension. The method can be used to drag out the implicit knowledge that is part of passage representations and schemas. The method can also be used to trace the process of constructing structures while passages are comprehended on-line. Our research has uncovered dozens of informative trends which have important implications. However, it is beyond the scope of this presentation to report many of the interesting observations (see Graesser, 1981). It suffices to say that the data are sufficiently rich and distinctive to discover new pieces to the puzzles of comprehension.

References

Graesser, A.C. Prose comprehension beyond the word. New York: Springer-Verlag, 1981.

Graesser, A.C., Robertson, S.P., & Anderson, P.A. Incorporating inferences into narrative representations: A study of how and why. Cognitive Psychology, 1981, 13, 1-26.

ROLE OF CONTEXT IN COGNITIVE DEVELOPMENT[1]

Gideon Carmi, Hebrew University, Jerusalem, Israel

(Theoretical Exposition and Preliminary On Experiments)

Purpose and Significance

The purpose of this study is to contribute both theoretically and practically within the area of cognitive development and conceptualization. We will discuss the effects of manipulating concept supporting and concept-distracting perceptual cues on performance in cognitive tasks - an area which has been virtually ignored by Piaget and his colleagues. We propose two main effects: 1) that the manipulation of perceptual cues will reveal the capability of both low and middle-class socio-economical-status (SES) children to reason concrete-operationally well before the ages posited by Piaget, or posited by investigators of cognitive functioning of low SES children, once and perceptual dressing of the standard tasks is made more digestible to these children, 2) that by exposing these children to an effectively spaced sequence of perceptually varying tasks which lead up to standard Piagetian or Piaget-like tasks, they will perform successfully also on the latter, again well before the posited age range for such performance. This is intended to serve as a model for accelerating cognitive functioning in general, whether at school, in everyday life or in vocational training.

From a theoretical standpoint, we propose to contribute thereby towards the crystallization of a model of cognitive development which integrates the "cognitive-change position" of Piaget and others, with the "perceptual-change position" of Odom and others (Odom, 1978). This model - which could be called a "cognitive-perceptual position" has been emerging out of field work in "Creative Maths and Science Teaching" with low SES children in Israel2 (1976-81), in Venezuela3 (1978-81), Costa Rica4 (1981) and Brasil5 (1981). This model takes account of the perceptual input as an additional dimension to cognitive processing, yet is closely and causally interwined with the latter. Our model is a continuous model from which Piaget's stage-like structures should, in principle, be derivable without further assumptions as a result of sudden overlaps of continuously changing regions. Piaget's experiments are seen as isolated points, each on another continuous dimension along which a "salience"-paramenter and changes. This model leaves ample room for cultural, social and other differences in cognitive development, especially in the attainment of stage-like behavioral criteria. The model also indicated the theoretic tools for explaining and predicting such differences.

Its major significance should be seen in its potential to provide a theoretical basis as well as practical tools for examining and accelerating cognitive and conceptual development, especially for disadvantaged children.

---

1. Supported by the Ford Foundation and by the Israeli Ministry of Education.

2. Project Petakh (PP), short for "development of thinking", directed by G. Carmi at the Department of Science Teaching at the Hebrew University, Jerusalem, and sponsored by the Israel Ministry of Education. This study is with 36 teachers in Jerusalem and with several replication groups elsewhere.

3. Ciencia Creativa (CC), (Creative Science), directed and implemented by CENAMEC (Centro Nacional para el Mejoramiento de la Ensenanza de la Ciencia), with the author as principal investigator. This is an adaptation of PP in Venezuela.

4. Ciencia Creativa (CC), directed and implemented by CEMEC (Centro para el Mejoramiento de la Ensenanza de la Ciencia of Costa Rica), with the author, G.C., as principal investigator. This is an adaptation of PP.

5. Ciencia Creativa (CC), pilot sponsored by CAPES, Ministry of Education, Brasil wth the author, G.C., as the principal investigator. This is an adaptation of PP.

CONCEPT FORMATION THROUGH THE INTERACTION OF MULTIPLE MODELS

Mark H. Burstein
Yale University

INTRODUCTION

This paper is about some of the processes involved in learning the fundamental concepts of a new domain. In particular, I have been studying how people learn some of the basic concepts used in computer programming, given little or no prior experience with the use of computers, or computer languages. The goal of the research is to build a computer model capable of learning such basic concepts.

The problem is that most current theories are either too weak or combinatorially explosive. Generalization techniques [4,6] do not construct radically new conceptual structures. Concept clustering techniques [2] require correlations over large sets of data. People learning what a computer variable is must learn something radically new to them, but with very limited experience. How can they do this?

The theory I am developing is based on the study of transcripts of 1-on-1 tutorial sessions with several subjects learning the programming language BASIC, collected over a period of several months. The subjects lacked any previous experience with computers, but even the youngest, Perry, age 10, had a number of preconceptions about what was to be learned, and a wealth of knowledge and experience in other domains which he used to provide correlates for experiences and observations in the new domain.

Learning to program a computer is a rather "formal" subject, requiring the development of expertise in symbolic manipulation. Thus one might readily expect prior knowledge of mathematics to be useful in learning to program. Indeed, our subjects often drew parallels to concepts in mathematics when learning programming concepts. Perhaps less expectedly, however, our subjects also invoked a great deal of common sense world knowledge when learning to program.

This common sense knowledge included such diverse areas as putting objects in boxes, moving pieces around on a game board, and writing on blackboards.

This work was supported in part by the Advanced Research Projects Agency of the Department of Defense, monitored by the Office of Naval Research under contract N00014-75-C-1111, and in part by the National Science Foundation under contract IST7918463.

It also included knowledge of English, and of how agents perform tasks requested of them. In this short paper, I hope to at least give some idea of how these other areas of expertise were used to form some very basic concepts of computer programming.

HOW ARE CONCEPTS IMPORTED?

One of the techniques a teacher can use to explain an unfamiliar concept is to use analogies or metaphors [5]. For example, a common way of introducing the concept of a computer variable is to make an analogy to boxes and their contents. One introductory BASIC manual phrased it this way:

> (1) To illustrate the concept of variable, imagine that there are 26 little boxes inside the computer. Each box can contain one number at any one time. [1]

Another text referred only to "locations" inside the computer, but illustrated the concept with diagrams showing contiguous rectangles containing numbers. Both of these descriptions suggest that essentially the same set of relations and operations that exist between containers and their physical contents can be applied to variables and their "contents".

However, this is not the only way to describe variables. Another common description is based on similarities to concepts in elementary algebra.

> (2) In ordinary algebra, letters of the alphabet are ofter used for terms that can take on varying numerical values. BASIC has a very nice, built-in algebraic feature that allows the user to assign numeric values to any letter of the alphabet... The user can set the value of A by typing A=10. [3]

While each of these models suggests a number of things that are true for variables, neither of them tells the whole truth and nothing but the truth. Although each analogical notion of variable presented suggests enough of the truth to act as an initial, working model of the functioning of variables in computer programs, they also suggest hypotheses which are either irrelevant or wrong, or both.

Since neither model is sufficient to give the learner an understanding of how to use variables, or how they will work in all situations, each model must be "debugged". [9] That is, having seen the metaphor, the student is presented with simple problems to be solved or new situations to be understood in the new domain, which he attempts to solve using hypotheses suggested by the metaphor. This leads to some successes, and a number of failures. Interestingly enough, the failures provide the most useful information about how the developing concept must be modified so that it agrees with first-hand experience. [7,8] The causes of the failure must be fixed must be fixed to produce useful expectations in the future.

USING MULTIPLE MODELS

One kind of modification that can be made to an inaccurate model is to add restrictions to prevent specific erroneous inferences. For example, there are no inherent restrictions on the number or kind of objects that can go in boxes. Variables, on the other hand, can hold only one thing at a time. Hence, a statement of that restriction must be added by the teacher, as in (1) above, or by the student after an error occurs.

A more important kind of modification of the developing concept is to introduce additional models or analogies. These additional analogies can be used to explain and/or to replace faulty inferences made by earlier versions of the new concept. The new analogies also allow additional inferences and expectations to be made.

For example, a second extremely common analogy for computer variables and computer memory is human memory. The following is an example of such a statement:

> (3) The computer "remembers" any values assigned to numerical variables. Of course, the values of the variables can be changed at any time -- simply do the variable = value command again. [3]

This metaphor for computer memory comes from the same text that contained the algebraic description of variables, (2), above. As we will see shortly, such secondary models can and do affect the hypotheses which can be generated.

These examples illustrate the use of multiple models in describing a new concept. This is not unusual. Rather, it may be the norm. Furthermore, the students will add metaphors and models of their own discovery, based on chance word choices, observed consistencies, and inferences they made.

A model of learning must explain how these analogical descriptions can be "combined" when forming new concepts.

GENERATING ERRONEOUS HYPOTHESES

Here are some examples of the kinds of errors and corrections that occur when a student has several models for the same concept. In the following examples, M is myself, and P is the student, Perry. Perry was initially given a short paragraph to read which said that the computer's memory used variables which were like boxes in which one could store numbers.

Perry produced the following summary of that description:

AN INITIAL HYPOTHESIS

P: Yea. This, the variable, is the box, and you tell it to... It's where he puts his memory and everything... and then you ask him to put it in there. You say the word (variable), and then he puts it in.
M: Right, ok. So suppose I wanted to tell him to remember the number 5 in variable X...
P: Oh. What do you mean X? In any box that he wants?

This shows that for Perry, variables were related to (human) memory, and that they were somehow like boxes which you put things in. His last statement also shows that he already had the concept of an algebraic variable, since here he thought X stood for the unknown. This fact will be important in other examples.

Using the box metaphor, Perry tried to investigate directly the class of objects that can be placed in the computer's "boxes". Since physical objects can certainly hold other physical objects, he thought that the boxes in the compter might be able to hold other "boxes".

STORING BOXES IN BOXES
-- another (partially) wrong hypothesis.

P: Can he store a box in a box? Let's say you tell him X=7+4, and then you make him another box and tell him to store X, the box X into the other box.
M: You can't put boxes inside boxes, but you can put what's inside them in another box.

Another error made in trying to specify what can go into the boxes comes from the use of a second metaphor. As this was also supposed to be a model of the computer's memory, Perry often generated hypotheses consistent with his knowledge of human memories. Since people, when asked to solve simple arithmetic problems, will generally remember both the question and the answer, Perry assumed that the computer would as well.

STORING THE PROBLEM WITH THE ANSWER

M: Suppose I wanted to add up 7 and 4 and store it in X.
P: How? You want to put 7+4 or the answer?
M: I want to put the answer in.
P: You want him to tell me the answer and then put it in?
M: Well, just put it in.
P: You write... X equals 11. You can write X=7+4
M: You can do that too. Try it.
P: (types X=7+4)
M: Ok. Now what's inside the box called X?
P: 7 plus 4, ... 11. You want me to ask him?
M: Yea, why don't you ask him?
P: (types PRINT X) oh. 11. It doesn't matter. If you want him to only answer 7+4, and not the answer, can it do that?
M: No, it only stores one number in each box.
P: He thinks you only want the answer.
M: It's not that, it's just that what you put in the box was not "7+4", but 11. It adds it up and then puts the answer in the box
P: Oh. It doesn't put 7+4 in the box.

The addition of semantic cues to the names of variables caused Perry to assume that the computer was even more like a human agent, with human-like memory capabilities. Here, he wonders whether the computer has the ability to querry it's "frame" for a person in several different ways.

THE COMPUTER AS HUMAN AGENT

M: I can say PERRY=10. How old are you?
P: 10
M: Ok, and now if I PRINT PERRY.
P: It will say 10
M: Right.
P: If you tell him... Let's say Perry is 10 years old. Can you say that? And then you ask him "How old is Perry?" or "Who is 10 years old?", will he answer you?

Finally, there is the case in which several of Perry's models make conflicting predictions, and he is asked to choose which one is correct. In this example, the two models are the box metaphor (assignment is like putting something in a box) and presumed similarities to the algebraic notion of equality (A=10 means A is the same as 10), which Perry inferred from the use of the equal sign to denote assignment in BASIC. If assignment is like moving something into a box, then moving something from one box to another should remove it from the first box. However, if assignment is like algebraic equality, then the statement Q=P should mean that the two variables have the same value.

After I had typed P=10 and then Q=P, (causing both Q and P to have value 10), the following dialogue occurred:

GENERATING CONFLICTING HYPOTHESES

M: So, what's in P now?
P: Oh. Nothing.
M: Nothing?
P: 10! and then Q is also.
M: What do you think it is? Is it nothing or 10?
P: Let's find out. First let's see...
M: Well, what do you think it is?
P: If you have two boxes, and you moved...you moved or it equals to? You moved what's in P to Q so there's nothing in it, or did you only put the same number in Q that's in P? I think it's 10.
M: You think it's 10?
P: Because you don't say that, um, move P at all... take P out. You only said that Q equals the same as P. So if it equals, it has to be 10, because if there's no 10 in it, Q wouldn't equal to it.

Clearly, there is great indecision here about which model is going to prove correct. Only after reviewing both choices does he finally chooses the one which makes the correct prediction.

This example clearly shows Perry operating with at least two distinct models of the what a variable is. Each model is capable of making a number of predictions which are useful. Each makes predictions which must be corrected. Thus, corrections are possible not only when a preferred model produces an error, but whenever two models contradict each other. In both cases, more questions are raised to guide inference experiments in the "model debugging" process.

In this example, Perry was forced to reiterate the two models so that he could reevaluate his choice, and find the one that seemed most consistent with the situation and his beliefs. In this case, the answer 10 seemed preferable due to the use of the phrase "Q equals P".

My goal is to develop a computer model capable of learning concepts radically different from those it already has, without requiring huge amounts of exemplary data. The technique suggested by the protocols is to maintain a set of models that are all partially useful, but which occasionally contradict each other. Much of the learning arises from the correction of the failed expectations of these models. However, these failures can occur even without the student "making a mistake". Contradictory expectations generated by competing models may provide an important alternative source of errors for the debugging process.

REFERENCES

[1] Albrecht, R., Finkel, L, and Brown, J.R. (1978). BASIC for Home Computers. John Wiley & Sons, Inc., NY.

[2] Dietterich, T.G. and Michalski, R. S. (1979). Learning and generalization of characteristic descriptions. Proceedings of IJCAI-79, Tokyo, Japan.

[3] Heiserman, D.L. (1981). Programming in BASIC for Personal Computers. Prentice-Hall, Inc., Englewood Cliffs, NJ.

[4] Kolodner, J.L. (1980). Retrieval and organizational strategies in a conceptual memory: A computer model. RR#187. Ph. D. Thesis, Yale University New Haven, CT.

[5] Lakoff, G. and Johnson, M. (1980). Metaphors We Live By. The University of Chicago Press, Chicago, IL.

[6] Lebowitz, M. (1980). Generalization and memory in an integrated understanding system. RR#186. Ph. D. Thesis, Yale University, New Haven, CT.

[7] Papert, S.A. (1972). Teaching children thinking. Programmed Learning and Educational Technology, Vol. 9, No. 5.

[8] Schank, R.C. (in press). Dynamic Memory: A theory of learning and reminding.

[9] Sussman, G.J. (1973). A computational model of skill acquisition. AI-TR-297. Ph.D. Thesis, MIT, Cambridge MA.

A Simulated World for Modeling Learning and Development

Pat Langley, David Nicholas, David Klahr, and Greg Hood[1]
Carnegie-Mellon University

1. Introduction

The creation of a complete intelligent system is a goal of many researchers in cognitive science and artificial intelligence. An elegant path to realizing that goal would be to let a system *develop* its intellect through interactions with the environment. Recent work in perceptual and motor skills has brought these hopes closer to fulfillment, but we are still many years from constructing systems that can adequately interact with the real world.

In this paper we describe a *simulated* world for developing and testing models of learning and development. Such a facility could play an important role in the emerging field of cognitive science, since it would encourage researchers to construct complete systems, and since it would provide a common ground on which competing theories might be compared. Below we summarize the aims of the project and the design criteria for the simulated environment. After this, we present an overview of the simulated world and of the sensory/effector interfaces through which model organisms may interact with it. We close with a brief discussion of the interface between the user and the environment.

2. Aims of the Research

The simulation system described here was conceived by, and for, cognitive scientists representing a variety of theoretical biases and paradigms. Its purpose is not to embody particular assumptions about complex information processing (human or otherwise), but rather to provide a medium for empirical research that can be shared by -- and tailored to the needs and goals of -- as wide a range of investigators as possible.

2.1. Focus on Learning and Development

The major goal of cognitive science is to understand the nature of intelligence. However, the knowledge and strategies responsible for intelligent behavior vary with time. Since any science searches for *invariant* regularities of behavior, this presents a difficult problem for our field. Langley & Simon (1981) have argued that we may find the desired invariants in a theory of learning and development, and preliminary steps towards a theory of the transition process have been taken by Klahr & Wallace (1976), among others.

Theories of learning and development can be characterized as having two problems to solve. First, they need to account for behavior at several different points in a developmental progression. Second, they must account for the mechanisms that enable the system to progress from state to state. The quality of the transitional theory is constrained by the quality of the theory that accounts for behavior at each state. One natural consequence of this argument is a focus on "expert" or "mature" performance: such models provide an ultimate target for the transitional processes, and would thus appear to have top priority in theory building efforts.

However, we believe that this emphasis has resulted in an unfortunate limitation. The contact between current computer simluation models and an empirical base is almost exclusively via adult performance measures, and often skilled adults, at that. To the best of our knowledge, none of the currently proposed mechanisms for taking a system from state N to state N + 1 could plausibly have brought the system to state N in the first place. That is, most models have little *developmental tractability*, although they may provide a reasonable account of the learning mechanism in an already well developed system.

In order to remedy this situation, we need to ask questions about infant systems, about the rudimentary encodings and internal representations, and about the innate kernal of self-modifying processes. We think it is important to attempt to fomulate developmental theories in which the ability to undergo self-modification in a plausibly supportive environment is the *primary* constraint on system design, with performance a problem to be solved within that constraint. (The converse is almost universally the case with modern theories of self-modification: performance models come first, with self-modification added on in ingenious ways.) The simluated world proposed here would facilitate such an effort. We envision our simulation system as a tool for investigating any of several domains of knowledge development, including form perception, object constancy, problem solving, or quantitative processes.

2.2. Constructing Complete Systems

A standard approach in science is to partition a phenomenon and study the pieces separately. In cognitive science, we find researchers specializing in language processing, problem solving, perception, and many other areas. And though this division of labor has clarified much about the components of intelligence, it has revealed little about their interaction. For example, researchers in language acquisition have separated that form of learning from concept formation and word learning, despite their strong interplay. Similarly, problem solving theorists have ignored perceptual and motor factors, though they may significantly influence the difficulty of a problem. We feel enough is understood of the components to initiate attempts to construct complete intelligent systems.

A second argument for creating complete systems comes from representational considerations. As long as one focuses on only an artificially bounded *subset* of behavior, the input of the system must be specified by the user. Thus, one might build a system with apparently general learning mechanisms, but which would learn only when presented with carefully hand crafted data. The construction of a complete system should guard against such subtle kludges, since information would be obtained through direct interaction with the environment or by inferences the system made itself.

2.3. Simulated Worlds vs. Real Worlds

The ideal complete intelligence would be a robot, with sensory abilities for perceiving the real world and motor abilities for affecting it. Unfortunately, we are still far from understanding perceptual and motor behavior in the necessary detail. This leads us to propose the less impressive but more manageable option of devising a simulated environment. The notion of a simulated world has its own attractions, including its relative independence of hardware and its transportability between sites, making it an ideal tool for cognitive scientists to employ in their model construction. It also eliminates the computational constraint that cognitive processing be done in real time.

3. Criteria for a Simulated Environment

In order for the simulated world to be useful as an experimental tool in cognitive science studies of learning and development, the following points were considered central design criteria:

- ***Independence and richness*** The world model must be truly separate from the organisms and their internal "models" of the world. Moreover, the world must be sufficiently rich and unpredictable, so that no organism can internalize a complete model of the world in its lifetime. In particular, the environment should be much richer than either Becker's (1970) grid universe or Winograd's (1972) blocks world. This is crucial, since the real world elicits qualitatively different behavior than would a completely-internalizable world (where table-lookup and formula-evaluation would suffice for perfect behavior).
- ***Extensibility and consistency*** -- The world must be extensible in terms of introducing arbitrary numbers of new objects and new organisms. However, physical laws must remain invariant.

[1] The ideas presented here evolved out of the interactive efforts of the CMU World Modelers Group, in which Jaime Carbonell has played a major role and which has included Hans Berliner, Greg Harris, Marty Herman, and Glenn Iba, in addition to the present authors.

- ***Cross-sensory correlation*** The world model should support multiple sensory media, and each organism can, in principle, be designed to take input from any subset of the sensory media. Cross-medium sensory correlations play a central role in theories of perceptual and cognitive development, and yet few computational studies of this phenomenon have appeared to date.
- ***Multi-level sensory interfaces*** -- In order to satisfy the objectives of different researchers exploiting the same world-model tool, interaction between the world model and the organisms should be mediated via sensory and effector interfaces, whose purpose is to provide the organism with data (or effector functions) specified at the desired level of abstraction.
- ***Synchronous processing*** -- The environment should not be forced to stop in order for the organisms to think at their leisure. Therefore, if the organisms are unable to process all incoming input, the focus-of-attention problem is introduced as an integral aspect of cognition.
- ***Communication among organisms*** The only communication possible is via the sensory and effector interfaces (by gestures, language, etc.), requiring organisms to pay attention in order for the communication to take place.

In formulating the above criteria we considered various issues we wished to research using the simulated-world environment. For instance, learning purposive action is a basic function in most higher-level organisms, including human "rational" thought. We wished to provide organisms with basic drives (e.g., hunger, curiosity, companionship of like organisms, etc.), basic actions, and learning mechanisms. Exactly what the starting point for learning processes ought to be is a matter or research and/or discretion according to the phenomenon one wishes to investigate.[2] For example, one type of purposive action is to learn subsumption goals (Wilensky, 1978); e.g., secure more of a resource (such as food) than presently required to satisfy an internal drive -- if past experience has shown that the drive will recur and finding the resource may be an uncertain or costly operation. Another is to posit intrinsic satisfaction from the simple, repeated execution of activities: e.g., Piaget's "circular reactions."

4. An Overview of the Simulated Environment

The simulated three-dimensional environment contains *objects* of two types: *primitive* (the building blocks of the physical domain) and *complex* (hierarchical structures, aggregates of primitive objects). Every physical structure, including the manifestations of organisms in the simulated world, is either a primitive or a complex object.

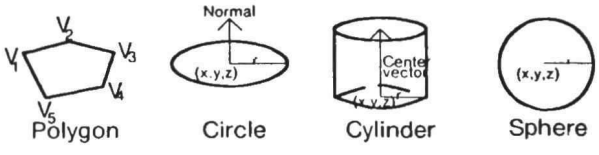


Figure 1. Primitive Objects

4.1. Primitive Objects

Figure 1 illustrates the four types of primitive objects (polygon, circle, cylinder, sphere) and the spatial parameters that must be defined for each. A polygon is specified by an ordered list of vertices, a circle by the coordinates of its center, a radius, and a normal vector, and so on. All primitive objects have an additional set of physical properties that must be specified: mass, center of mass, velocity, angular velocity, elasticity, static and dynamic coefficients of friction, temperature, taste, color, and texture.

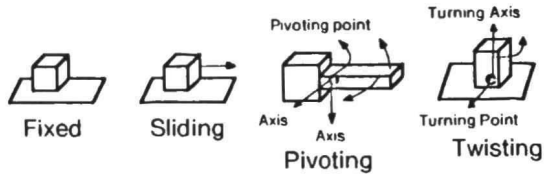


Figure 2. Joint types used in complex objects

4.2. Complex Objects

The joining of primitive objects to one another, in any of the several ways shown in figure 2 (joints may be fixed, sliding, pivoting, or twisting), produces a complex object -- an organized structural hierarchy within which properties may be shared and which may be acted upon by the physical laws of the environment as a single unit. A joint may be broken, and the complex object separated into two (reorganized) hierarchies, by the application of force in excess of that joint's prespecified *stress limit*. A fifth, or "virtual," joint type may also be employed, when two objects are touching and not moving relative to one another (i.e., in static contact).

Consider the example of a table with a coffee cup resting on it. A typical table has four legs and a top: four primitive objects (long, thin cylinders) connected by fixed joints to a fifth (a circular or polygonal plane). The cup may be broken down into its body and a handle: the former a hollow cylinder closed at one end by a circle of equal radius, and the latter three more cylinders connected to resemble three sides of a square. A virtual joint connects the two complex objects, so that if a gentle force is applied to the table it will take the cup with it when it moves (depending, of course, on the exact magnitude and direction of the force, the frictional coefficients of the table top and the bottom of the cup, the inertia of the cup, etc.).

4.3. Physical Laws

The simulated environment must be "updated" at the start of each quantum of time. The positions of objects need to be adjusted, based on previous velocities and the application of forces; stresses on joints must be computed and, if necessary, objects fragmented. In short, changes that have resulted from the influences of objects (including the model organisms) upon one another -- or from some alteration introduced from without, by the user -- must be incorporated into the description of the current "state-of-the-world." This is accomplished by applying (recursively, to complex objects and their components) a set of simplified physical laws: rules for the modeling of relatively gross interactions in the environment.

5. Sensory and Effector Interfaces

Organisms must interact with the simulated environment in a manageable way that can be tailored to the needs of the researcher. As figure 3 illustrates, the sensory and effector interfaces (also called sensory-motor interfaces, or S-MI's) are independent of both the organisms and the external world model. The primary reason for this decision is that different researchers should be allowed to define the level of abstraction of the sensory information detected by each class of organism. Hence, a researcher interested in language acquisition can "plug in" a symbolic visual interface that provides the names and locations of objects in the visual field of the organism (in order to relate objects and actions to words and sentences). However, a researcher interested in perceptual or motor learning can substitute a lower-level interface that provides only collections of features for visible objects -- or yet lower-level visual data "decompiled" from the world model by the sensory interface. The same design principle holds for effector interfaces.

---

[2] One must be careful in researching learning by simulating autonomous organisms not to fall into the self-organizing-system fallacy. Early AI research in learning postulated extremely simple learning mechanisms and a virtual *tabula rasa* with respect to real world knowledge, with the expectation that such a system could organize itself into a thinking entity. The only parallel of such a "magical" feat in the real world is the evolutionary process, but this process required billions of years, billions of generations of organisms, and millions of individual organisms per generation. Our world model is designed with the objective of modeling developmental learning -- where a single organism can learn purposive action in a fraction of its lifetime. As such, it requires a non-zero starting point with some innate abilities, a high communication band width with the external world, and built in drives to focus its attention and guide its behavior (at least initially).

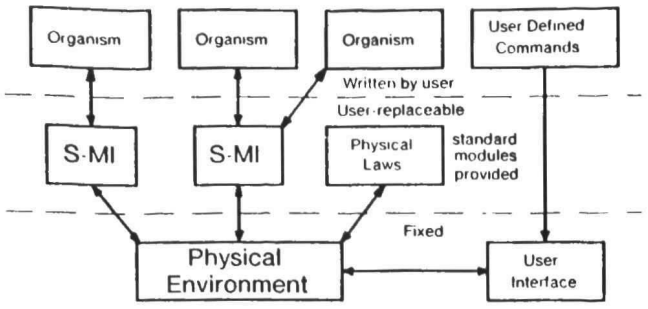


Figure 3. Modules of the System

We believe that the sensory and effector interfaces should themselves be subjects of research. As an example of the problems that arise in designing a sensory interface, consider the fact that human beings can tell with precision the location of nearby objects, but can only place approximate values on the absolute location of more distant objects. This phenomenon seems crucial for some forms of generalization. For example, it does not matter if a predator is 1001 or 1003 feet away the organism should flee; however, in reaching for an object that is 1 or 3 feet away, different actions are called for. Therefore, the perceptual mechanism that accentuates the latter difference, while glossing over the former, appears to be a desirable feature.

6. The User Interface

To exploit the advantages of a simulated world it is essential to have a powerful user interface Of major importance is the ability to view the world as an observer within the world would see it. In our implementation, the user may create multiple windows, each of which is a perspective drawing of the world from a different viewpoint These viewpoints may be fixed, or they may be bound to the "eyes" of particular organisms within the world so that the window accurately portrays the image seen by the organism. A graphical display also allows the user to easily construct new objects and organisms differing in physical characteristics.

In addition to having commands for controlling the windows, the user must also be able to control the motor and verbal behavior of certain organisms that function as teachers (of other organisms). The interface should allow the user to specify actions as high level commands instead of cumbersome primitives (such as forces to be applied at certain joints). Simulation of the world also enables a detailed record of events to be kept for later examination by the investigator. Assuming the cognitive states of the organisms are saved periodically, such a record could be backed up to a certain point in time and the simulation restarted, providing a valuable opportunity to explore alternative courses of action.

7. References

Becker, Joseph D. An information-processing model of intermediate-level cognition. Stanford University Department of Computer Science AIM-119, May, 1970.

Klahr, David & Wallace, J. G. Cognitive Development: An Information-Processing View. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1976.

Langley, Pat & Simon, Herbert A. The central role of learning in cognition. In J. R. Anderson (ed.), Cognitive Skills and Their Acquisition. Hillsdale, N.J.. Lawrence Erlbaum Associates, 1981.

Wilensky, Robert. Understanding goal-based stories. Yale University Department of Computer Science Research Report 140, September, 1978.

Winograd, Terry. Understanding Natural Language. New York: Academic Press, 1972.

Why Do We Do What We Do?

James A. Galambos
and
John B. Black

Yale University

Abstract

We examined the reasons people give for the actions in routine events and classified them as action enablement, main goal satisfaction, or external. We argue that the external reasons indicate that some actions in events are stored at more general levels in memory than the specific event schemas.

There has been a recently rekindled interest in the relationships between the goals of common activities and the structure of our knowledge of these. A number of researchers (Schank & Abelson 1977, Wilensky 1978, Graesser 1978, Lichtenstein & Brewer 1980) have argued that the goals of familiar events may have stereotypic plan paths associated with them. In this paper we examine the goals and plans for a large number of common activities in the hope of isolating some of the parameters which effect their knowledge representation.

We chose thirty common events for analysis. These included events such as the familiar GOING TO RESTAURANTS activity, as well as SHOPPING FOR GROCERIES, GOING TO MOVIES, CHANGING A FLAT TIRE, BREWING SOME TEA, WASHING CLOTHES, WRITING A LETTER, etc. For each event we picked twelve component actions which described the event from beginning to end. We chose them with the constraint that they not overlap temporally in the performance of the event and that they be at about the same level of description. For example, some of the component actions for the grocery shopping event were MAKE A LIST, GET A CART, LOAD THE CART, GO TO CHECKOUT, etc. This level of description was felt to be neither too molar (containing a number of discernable actions) nor too molecular (specifying fine-grained motor movements).

These same stimulus materials have been used in a number of other experiments investigating the structure of memory for events. In the course of those studies, a number of norms were collected including the <u>importance</u> or centrality of the component actions for the event, the frequency of performing the action when doing the event, the sequential order of the actions and finally the frequency of being in the event when performing the action. This last measure I have called the <u>distinctiveness</u> of the action to the event. For example the component action EAT THE MEAL is not very distinctive to the restaurant event since the action of eating is done in many other situations. In this example the distinctiveness of the action and its importance to the event are orthogonal in that eating is highly important to the event but not highly distinctive. In contrast the action SEE HEAD WAITER is highly distinctive to the restaurant event because it is done almost exclusively in that situation. Distinctiveness of an action can be seen as a measure of the extent to which that action has an independent existence outside the context of the event. This factor will figure in our discussion of the results.

The present study involved presenting subjects with event-action pairs and asking them to write down their reasons for performing the action in the context of the event. The original intent was to use the results to develop superordination and subordination relations among the actions as has been done in studies by Graesser and Lichtenstein & Brewer. There are a number of methodological differences between this study and the others. Perhaps the most salient of these is the absence of a particular instantiation of the events and their components via the presentation of a videotape or story depicting the event. Our subjects were asked to rely on their

knowledge of the events to help them provide their reasons for the actions. When subjects are presented with an instance of an event, they understand it by interfacing their general knowledge of that type of event with the explicit details provided by the story. The results of testing memory performance for that story must be interpreted as some function which includes contributions from both the specific details and the pre-existing knowledge base. A further source of potential variance arises from reconstructive strategies which occur at the time of testing. Most research in this area has attempted to minimize the effect of the specific detail in the story by devising stimulus materials which are as dull and boring as possible. This is cited as a methodological virtue because dullness is thought to indicate that the story matches the most typical or least deviant plan path through the actions in the event. While this general technique had yielded a great deal of insight into comprehension and memory for text, it is important to note that the knowledge base for an event prior to instantiation may differ significantly from the representation that results from its application to the task of understanding a particular instance.

Specifically, our underlying knowledge of common events is flexible enough to permit a wide variety of different realizations of those events. For instance, while the restaurant event may often occur in contexts where the primary goal is to satisfy hunger, it also occurs in the service of other overriding goals such as to celebrate some notable happening, or to conduct some business, or to fulfill a desire for an evening's entertainment. Performance of the event under these differing circumstances may alter the salience of certain actions in the event by changing the internal goal structure and consequently the plan paths connecting the component actions.

By analyzing the reasons that subjects give for the actions it is possible to outline the range of flexibility in the representation of events as goal directed. The type of reason given for an action can be an index for the sort of goal that governs its presence in the event. Furthermore when subjects give more than one type of reason or do not agree highly as to the the best reason then it is often possible to glimpse some of the parallel goals underlying the event.

A first step in this analysis is the categorization of the reasons into general types. Clearly this analysis has a number of methodological difficulties. This free generation paradigm tends to lead to idiosyncrasies in the responses obtained. However in an exploratory experiment such as this such lack of constraint can be considered a virtue. It is often the response which is the outlier in a frequency distribution which can provide an important insight. With this in mind we attempted to organize the reasons into three types. The first we call the <u>action enablement</u> type. It is this type of reason that Graesser and Lichtenstein & Brewer examine and is the most common type. This type of reason mentions immediately subsequent actions in the event. For instance in the CHANGING A FLAT event the the most common reason for the action of GETTING THE JACK mentions the subsequent actions of POSITIONING THE JACK and RAISING THE CAR. There are different types of enablement which when used to construct internal goal hierarchies result in different kinds of structures. Some of these enablement relations reflect or perhaps underly a fairly strict temporal sequence of actions. Discontinuities between the action sequences and the enablement structures often reflect segmentations that can be considered as the <u>scenes</u> of the event (Schank & Abelson, 1977).

The second type of reason we will consider is often given for the actions that immediately precede these points of discontinuity. Our second type of reason we call <u>main goal satisfaction</u>. A reason of this type involves mentioning the main goal of the event. This is often accomplished by merely restating the name of the event. For instance, in the WASHING CLOTHES event the segment which includes the loading of the washer and putting in the soap

appears to conclude with the action TURN ON WASHER. The reasons given for this action are "to wash the clothes" or "to get clothes clean". There is little mention of reason concerned with the enablement of the subsequent actions involved in drying the clothes. While this result is useful in specifying the relations of lower level event goals to the main goal of the event it also poses a problem for those who attempt to account for the serial order of the action on the basis of goal structure. The problem is that there are rather glaring gaps in the representation which must be bridged by some means other than internal plan paths. It is however outside the scope of this paper to consider possible ways to close these gaps. There are many instances of reasons of this second type which are given for actions that are not at scene boundaries. The presence of these further argue for a flexibility in the event representation whereby actions may or may not be dominated by internal goals depending on the particular instantiations of the event.

The final member of our typology of reasons is called the external variety. External reasons are those which mention very high level goals such as preservation of health, cleanliness, avoidance of danger or legal and social punishment, and maximization of pleasure or enjoyment. Some of these reasons are related to goals for which the event itself functions as fulfillment. Others are general modes of operation such as to preserve money. For example, in the GROCERY SHOOPING event the reasons for the action CHECK THE PRICE were "to save money for other things", or "to avoid being cheated." External reasons often tended to be the outliers we spoke of earlier. It is these actions which were often more general in the sense that they could occur in other events. Furthermore since the external reasons provided points of connection with more general goals they often offered a delimitation of the set of occasions for the performance of the event. By a more detailed examination of the interaction of the kinds of high level goals which could govern the entry into the event with the internal goal structure it may be possible to specify a range of acceptable variation in the types of stories which count as instances of the event. This in turn may provide a way to constrain the representation of events so as to attain the desired flexibility without losing the important aspect of stereotypy which binds different instances of the same event.

Schank (1980) proposed a theory which satisfies these two criteria. One of the important points in this model is the restriction of the role of the script representation. These structures have been stripped of any information which can be generalized out of the script and stored at higher levels in memory. In our terms this would mean that the event level of representation would contain only those actions and goal relations which were incapable of of directly fulfilling higher level goals or those which never occur except in the context of a particular event. Although it remains to be worked out in detail there is important support for this view which arises from the results of our analysis of reasons into the three types. It appears that those actions which received reasons of the external type are just those which are susceptible of generalization. Further support for this claim comes from looking at the relation of the distinctiveness rating of an action with the type of reason given for it. It turns out that highly distinctive actions are those which receive exclusively internal reasons and the less distinctive actions are those which have reasons which refer to higher level goals. Recall that the distinctiveness factor indicates the extent to which an action exists outside of a particular event. Thus the correlation of distinctiveness and type of reason provides evidence for the concept of generalizing information out of the script representation. Furthermore a careful analysis of the types of reasons given for actions may lead to a principled way to carry out this project.

In this paper we have attempted to argue for the importance of recognizing that the representation of common events must provide for a wide range of

variability. Indications of this range can be obtained by examining the various purposes for the component actions. Our rough taxonomy of reasons has yielded some evidence in support of a model of memory representation which permits such variability. While there are still a number of important issues which must be addressed, it appears that there are a number of theoretical benefits for considering why we do what we do.

Graesser, A.C. How to catch a fish: The memory and representation of common procedures. Discourse Processes, 1978, 1, 72-89.

Lichtenstein, E.H. & Brewer, W.F. Memory for goal directed events. Cognitive Psychology, 1980, 12, 412-445.

Schank, R.C. Language and memory. Cognitive Science, 1980, 4, 243-284.

Schank, R.C., & Abelson, R.P. Scripts, plans, goals, and understanding. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Wilensky, R. Understanding goal-based stories. Unpublished doctoral dissertation, Yale University, 1978.

An Activation-Trigger-Schema Model For the Simulation of Skilled Typing

David E. Rumelhart
Donald A. Norman
Department of Psychology
and
Program in Cognitive Science
Center for Human Information Processing
University of California, San Diego
La Jolla, California 92093

We thank Eileen Conway, Donald Gentner, Jonathan Grudin, Geoffroy Hinton, Paul Rosenbloom, and Craig Will for their assistance and many discussions with us on the nature of the typing data, their work in collecting and interpreting typing errors, and their discussions on the underlying response mechanisms. Gentner has provided a large set of keypress reaction time data. This ongoing research and the several large corpora of data on typing performance have been of considerable assistance to us in the preparation of this paper. (The several studies will be published as separate research reports.)

Research support was provided by the Office of Naval Research and the Naval Air Development Center under contract N00014-79-C-0323.

Requests for reprints should be sent to David E. Rumelhart, Program in Cognitive Science C-015; University of California, San Diego; La Jolla, California, 92093, USA.

Abstract

We review the major phenomena of skilled typing and propose a model of the control of the hands and fingers during typing. The model is based upon an Activation-Trigger-Schema system in which a hierarchical structure of schemata directs the selection of the letters to be typed and, then, controls the hand and finger movements by a cooperative, relaxation algorithm. The interactions of the patterns of activation and inhibition among the schemata determine the temporal ordering for launching the keystrokes. To account for the phenomena of doubling errors, the model has only "type" schemata -- no "token" schemata -- with only a weak binding between the special schema that signals a doubling and its argument. The model exists as a working computer simulation and produces an output display of the hands and fingers moving over the keyboard and it reproduces some of the major phenomena of typing, including the interkeypress latency times, the pattern of transposition errors found in skilled typists, and doubling errors. Although the model is clearly inadequate or wrong in some of its features and assumptions, it serves as a useful first approximation for the understanding of skilled typing.

The Basic Phenomena of Typing

The fundamental phenomena fall into three categories: those involving timing of keystrokes, those involving errors, and those involving the general organization of the typing process. In this paper we simply list the phenomena. In a larger version of this paper, the individual phenomena are discussed and illustrated in detail (Rumelhart & Norman, 1981).

I: The timing of keystrokes

A. People can type very quickly.
B. Cross hand interstroke intervals are shorter than those within hands.
C. Within hand interstroke intervals appear to be a function of the reach from one hand to the next.
D. The time for a particular interstroke interval can depend on the context in which it occurs.
E. There is a negative correlation between the intervals on successive strokes--especially when the alternate strokes occur on alternate hands.

II: Pattern of Errors

A. Transposition errors
B. Doubling errors
C. Alternation reversal errors
D. Homologous errors
E. Capture errors
F. Omission errors
G. Misstrokes

III: The general organization of typing

A. Skilled typists move their hands towards the keys in parallel
B. The units of typing seem to be largely at the word level or smaller
C. Sequences involving cross hand strokes seem to take longer to program than those involving only within hand strokes

A Model of Typing

We have constructed a model that has the following properties:

1. control of action sequences by means of schemata;
2. selection of appropriate motor schemata through a combination of activation value and triggering condition;
3. the representation of letter typing by means of a *pure type theory* (i.e., one with no type-token distinction).
4. the need for distributed (local) rather than concentrated (central) control of movement.

The basic framework that we follow is called an Activation Triggered Schema system (ATS). The model consists of a set of schemata, each with activation values. A schema has an activation value that reflects the total amount of excitation that it has received. When appropriate conditions have been satisfied, a schema may be "triggered," at which time its procedures become operative and control whatever operations they specify.

Figure 1 illustrates the basic structure of the model. The model incorporates the ATS system plus specific control mechanisms for the activations and selection of particular hand and finger movements. The input of the model is a string of characters that constitute the text to be typed. The output is a sequence of finger movements, either displayed on a visual computer-controlled display as the movement of the hands and fingers over a typewriter keyboard, or as a series of coordinate locations for the relevant body parts.

Figure 2 illustrates the basic assumptions of the activation process using the word **very** as an example. First, the schema for the word is activated by the perceptual system and parser. This, in turn, activates each of the child schemata for keypresses. Each keypress schema specifies the target position, with position encoded in terms of a keyboard centered coordinate system. These target positions are sent to the response system which then must configure the palm and finger positions properly. Each keypress schema inhibits the schemata that follow it. This means that proper temporal ordering of the keypress schemata is given by the ordering of the activation values. In addition, the activation values are noisy, which leads to occasional errors.

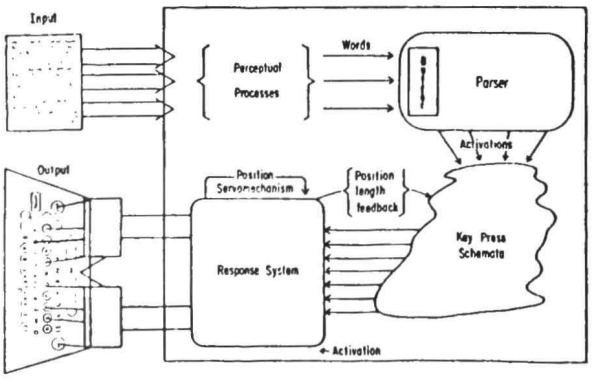
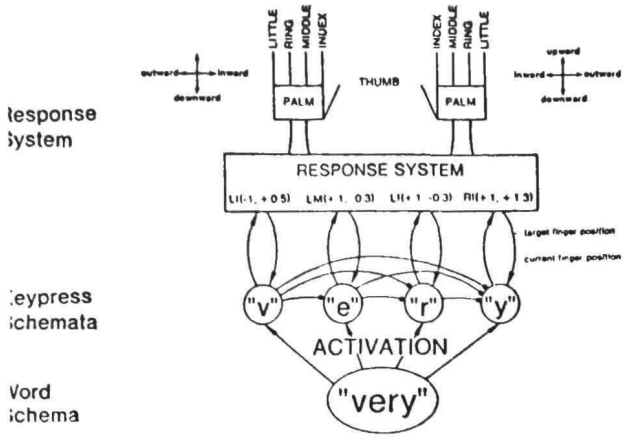


Figure 1. The information processing system involved in typing.



igure 2. The interaction of activations when the word **very** is to : typed. Inhibition is shown by the lines with solid circles at their rmination.

The response system feeds back information to the keypress system about the current location of the fingers. Whenever the current finger position is within some criterion distance of its target position, and the relevant schema is the one most highly activated, then the triggering conditions are satisfied and the actual keystroke is launched.

Repeated letters imply that there are no token schemata.

A doubling error occurs when a word that contains a doubled-letter is typed, but with the wrong letter doubled: **bokk** instead of **book**. An alternation error is similar: **thses** instead of **these**. The doubling error has two major implications. First, it implies that there are special schemata that signal the existence of doubled letters, and that occasionally these schemata get applied to the wrong letters; the binding between the doubling schema and its argument is weak. Second, the doubling error implies that each letter can only have a single keypress schema; only "types," no "tokens." In the model, the arguments are not bound to the schemata, but are established via activation values. After a keystroke schema has been triggered (and after the resulting launch of the keystroke), this keypress schema can become "bound" to a doubling schema if one exists with a higher activation value than its own. Because activation values are noisy, occasionally this leads to errors in the linking of keypress schemata to a doubling schema. (The existence of alternation errors leads to essentially the same conclusion.)

Movement

In the model, each active schema pushes its relevant hand and finger toward its desired key at the same time, and the final overall configuration is determined by the competition among these forces. Each schema pushes with a force proportional to its activation level. As a result, the forces are weighted so as to cause the letter schema that is next in line to be typed to approach its key most quickly. The actual location of each finger is determined by the sum of the extensions of the finger and the hand. To type a particular typewriter key, *it is only necessary that the end position be correct.* The endpoint configuration is reached through an iterative relaxation process that only involves local computation. Because of the unequal weighting of activations, the process will eventually cause the most highly activated schema to move its finger-palm configuration to within a criterion distance from its target key, satisfying the trigger conditions and launching the keystroke.

Appraisal

In order to evaluate the model, we gave it a text of slightly over 2,000 words to type. The pattern of keystrokes and times were collected from the simulation and analyzed in exactly the same fashion as we had analyzed the data from our human subject. Overall, the fit of the model to the many phenomena of typing is good. Detailed analyses of the model performance indicate that the simulation results do show about the right pattern of interstroke intervals. Moreover, the correlation of model times with actual typing times for the 66 most common bigrams from our data with the data of 6 subjects yields an overall correlation between the model and the averaged data of about 0.86. The fit is not bad, however, the model clearly does not account for all that is happening. (Data for five of these subjects were collected by Donald Gentner.)

The model does produce errors in typing, the most important ones being *transposition errors, doubling errors, and misstrokes.* The proportion of errors in the model is determined by the amount of noise in the activation levels. In our subject's data we observed transpositions at about a rate of 1 for every 1800 keystrokes. We adjusted the noise level to yield errors at a rate of about 1 for every 30 keystrokes. Despite the large difference in rate, the basic pattern of errors is similar. For example, a large majority (76 percent) of the transpositions in the simulation occur across hands. This is about the same as our subject and comparable to values reported for others. At the level of noise employed in the simulation, 17 doubling errors were generated.

Conclusion

We have constructed a working computer simulation of a model that captures the appropriate spirit of the phenomena observed with human typists, although it does not yet offer a complete account of the typing process. Despite the lack of an internal clock or "metronome" for timing, the model provides a reasonably good account of the timing patterns observed among skilled typists, including the prediction of negative correlations among successive keystrokes, a characteristic of metronome models. In similar fashion, there are no specific context dependencies built into the model and yet the time that it takes to strike keys depends upon the context in which they occur. We have no specific stored timing patterns for specific words, yet the model predicts that words have characteristic time profiles. We have no specific mechanism for transposition errors, yet our model generates the correct types of transposition errors. Moreover, the co-ordinative structure assumed within the model yields a qualitative emulation of the pattern of overlapping movements shown in a high speed film of a typist.

A number of conclusions can be drawn from our studies. First, the existence of doubling errors strongly implies the existence of a pure "type" representation of the keyboard schemata, with their arguments only loosely bound. Second, the nature of the skill requires simultaneous, parallel control of the fingers and hands, and this requires some form of negotiation process to turn the potentially competitive movements into cooperative ones. The degrees of freedom problem can be turned into a degrees of freedom virtue. Third, the model must incorporate the entire environment within which the typist operates, from the reading of the text, to the cognitive and motor control systems, to the shapes and mechanical characteristics of the hands, finger, and keyboard. Indeed, some of the limitations of the current model may really result from limitations of how well we dealt with the environment surrounding the control processes. Perhaps the central conclusion to be drawn from our analysis of typing deals with the nature of skilled motor co-ordination. We propose that the motor control system carries out its computations relatively locally and in parallel. We presume that such a conclusion will be proven for all skills involving high speed performance.

References

Gentner, D. R., Grudin, J., & Conway, E. *Finger movements in transcription typing.* San Diego, California: Center for Human Information Processing, University of California, San Diego. Technical Report 8001. May, 1980.

Rumelhart, D. E., & Norman, D. A. *Simulating a Skilled Typist: A Study of Skilled Cognitive-Motor Performance.* San Diego, California: Center for Human Information Processing, University of California, San Diego. Technical Report, 1981.

Toward a Formal Theory of Human Plausible Reasoning

Allan Collins
Bolt Beranek and Newman Inc.

Ryszard S. Michalski
University of Illinois
Champaign-Urbana

The talk will describe our attempts to formulate a theory of human plausible reasoning based on analysis of people's answers to a large number of everyday questions (Collins 1978a, 1978b). The formalization generates a large number of plausible inference types from a small set of basic elements.

The basic elements in the theory include:

arguments	v_1, v_2, etc.
descriptors	a_1, a_2, etc.
references	c_1, c_2, etc.
terms	$a_1(v_1)$, $a_2(v_2)$, etc.
statements	$a_1(v_1) = c_1 : \gamma$
mutual dependencies between terms	$a_1(v_1) \overset{\alpha\beta}{\longleftrightarrow} a_2(v_2)$
mutual dependencies between statements	$a_1(v_1)=c_1 \overset{\alpha\beta}{\longleftrightarrow} a_2(v_2)=c_2$

Mutual dependencies are bidirectional reflecting people's functional knowledge, as between say the likelihood of a person having mumps and also having fever. In these expressions α, β, and γ are certainty parameters: γ reflects the degree to which a person thinks a statement is true, α reflects the degree of certainty about the right hand term in a mutual dependency given that the left hand term is true, and β the certainty in the reverse direction.

There are also four operators that occur in the rules of inference: generalization, specification, similarity, and dissimilarity (which is the negative operator in the system). These operators are designated by a "rel" in the rules of inference.

The inference rules in the system as developed so far include: descriptor transforms and reference transforms on both statements and mutual dependencies, and derivations from relations between arguments and from relations between terms in both statements and mutual dependencies. Attribution inferences, which are common in human reasoning, are an elaboration of these basic inference types. The system does not yet include the rules for induction of statements and mutual dependencies, nor the various meta-inferences in the theory (Collins, 1978b).

Given that "rel" can be realized in four different forms, this system generates 64 different one-step inferences, and a very large set of two-step inferences. Many common human inferences are two-step inferences in the system, as for example the functional analogy described in earlier papers (Collins et al., 1975; Collins, 1978a, 1978b). We will try to show how a variety of human protocols can be accounted for, given such a system of plausible reasoning.

References

Collins, A. Fragments of a theory of human plausible reasoning. In D. L. Waltz (Ed.) Theoretical issues in natural language processing, Vol. 2. Champaign, Ill.: University of Illinois, 1978a.

Collins, A. Human plausible reasoning. Cambridge, Mass.: Bolt Beranek and Newman Inc. Report No. 3810, 1978b.

Collins, A., Warnock, E., Aiello, N., & Miller, M. Reasoning from incomplete knowledge. In D. Bobrow & A. Collins (Eds.) Representation and understanding. New York: Academic Press, 1975.

Skills, Learning and Parallelism

Aaron Sloman,
Cognitive Studies Programme,
University of Sussex,
Brighton, BN1 9 QN,
England.

The distinction between compiled and interpreted programs plays an important role in computer science and may be essential for understanding intelligent systems. For instance programs in a high-level language tend to have a much clearer structure than the machine code compiled equivalent, and are therefore more easily synthesised, debugged and modified. Interpreted languages make it unnecessary to have both representations. Further, if the interpreter is itself an interpreted program it can be modified during the course of execution, for instance to enhance the semantics of the language it is interpreting. (This sort of possibility vitiates many of the arguments in Fodor (1975) which assume that all programs are compiled into a low level machine code, whose interpreter never changes).

People who learn about the compiled/interpreted distinction frequently reinvent the idea that the development of skills in human beings may be a process in which programs are first synthesised in an interpreted language, then later translated into a compiled form. The latter is thought to explain many features of skilled performance, for instance, the speed, the difficulty of monitoring individual steps, the difficulty of interrupting, starting or resuming execution at arbitrary desired locations, the difficulty of modifying a skill, the fact that performance is often unconscious after the skill has been developed, and so on. On this model, the old jokes about centipedes being unable to walk, or birds to fly, if they think about how they do it, might be related to the impossibility of using the original interpreter after a program has been compiled into a lower level language.

Despite the attractions of this theory I suspect that a different model is required. In chapter 8 of Sloman (1975) I drew attention to familiar facts about children counting which suggest that instead of using a single program interleaving the production of a new numeral and pointing at a new object, they run two processes in parallel, using a third process to monitor them and keep them in step, or abort them if they get too far out of step. If children used a single serial program, repeating the steps

SAY NEXT NUMBER; POINT AT NEXT OBJECT;

within some kind of loop, then it would be impossible to get out of step. But they do, and sometimes spontaneously correct themselves. Adults performing some tasks requiring two sequences of actions to be synchronised, for instance playing a musical instrument with two hands, may experience similar problems.

The ability to run a program in parallel with others, using a third process to achieve synchronisation could be a powerful source of new skills. For instance, it would not be necessary to write a new program interleaving the steps of two old ones, as is required in conventional programming languages. Provided both programs are initially represented in a form which permits synchronisation with messages from other processes, it becomes possible to synthesise a new skill simply by running the two old programs in step. It may be necessary to develop new perceptual skills to check whether they are in step or not, but that sort of skill would be required in any case for monitoring a single serial program. Similarly, instead of re-writing a program to cope with different stopping conditions, the same program could be executed and interrupted by different external monitors: for instance counting all the buttons, counting out buttons till there's one for each button-hole, couting out five buttons.

If programs are to be run in parallel this can be done either by time-sharing a single processor, or by using a network of processors which can work in parallel. In principle the two are equivalent, though time sharing one processor raises many difficulties if each of the separate processes has its own requirements concerning speed of execution, synchronisation etc. Further, there is plenty of evidence that human and animal brains consist of many units which can do things in parallel. It is therefore most likely that if processes do run in parallel as suggested above, then they probably run on different processors.

This immediately suggests the possibility that different processors may have different computational resources. For instance they may vary in speed, or memory capacity. More importantly, they may vary in the extent to which they have capability to run programs or the extent to which they have access to mechanisms required for synthesising procedures, monitoring them, debugging them, interrupting and restarting them, relating execution steps to goals and percepts, and so on.

Thus there might be some processors with all the facilities required for developing programs, and other processors capable only of running the programs. If fully developed and tested programs produced by the former are shippped out to the latter processors for execution, then this could produce the kinds of phenomena, mentioned above which suggest to many people that compilation has occurred. By contrast, our model does not require a change of representation, merely a change to a different interpreter. The first process might develop the general structure of a skill or ability, perhaps leaving some of the fine tuning, adjustment of parameters and thresholds, etc., to be done at lower levels while the program is run by a different machine.

A theory along these lines could explain how many skills (e.g. musical performance) might be learnt by learning various subskills which are subsequently put together. The synchronisation of two old skills might involve the development of a new third skill, which will run in parallel with them. (Try opening and shutting your mouth and your fist repeatedly in time. Then try doing it out of phase.) More complex skills might involve an extended hierarchy of sub-processes some of which control others. Some sort of synchronisation between largely independent processes is in any case required for co-ordinating visual perception with movement of limbs.

There are different ways in which synchronisation might be achieved. The difficulty of playing a piano piece where the left and right hand use different beats, suggests that sometimes the co-ordination of two or more low-level machines requires synchronisation signals linked to suitable points in the programs. Synchronisation could make use of global timing signals, shared between all processes. Alternatively, different

groups of processes might use their own synchronisation signals. (The former would limit the number of different tasks requiring different rhythmic patterns which could be performed in parallel.) Further, some kinds of synchronisation might use a sort of variable representation of speed (like a throttle), as is suggested by the co-ordination of complex dance movements or the hand and foot movements needed to drive a car.

It is possible that other things besides timing can be co-ordinated. For instance in playing music with two hands, phrasing, stress and volume can be co-ordinated, and the same piece may be played with different superimposed 'expression', suggesting that there is a supervisory program which controls the way the sub-programs are executed. So besides timing, it seems that at least amplitudes and smoothness of execution can be externally controlled.

If complex actions involve many different processes running in parallel, then interrupting and re-organising the processes may be a very complex matter. Such disturbances seem to play a role in some emotional states for instance when you lose your balance, or are startled by a face seen suddenly at a window (Sloman 1981).

There are at least two different ways in which a program might be "shipped out" to a lower level process after being synthesised by a "central" processor. The whole program might be copied into the memory of the new processor. Alternatively, there might be common access to some memory, with new processors being told to get their instructions from the very same data-structure built by the program-synthesiser. The former might be more suitable where there is no shortage of processors or memory space, or where there is a shortage of rapid access communication paths. The latter might be appropriate where processors have to be re-used for different purposes, and where subsequent modifications to the program, achieved by the higher-level machine, should be immediately available to the lower levels.

There are many problems and gaps in this theory sketch, including unknown trade-offs. Is there only one program-synthesising machine, or are there several, allowing more than one new skill to be learnt at a time? (E.g. learning a new poem at the same time as learning a new scale on the piano? Learning the words of a song at the same time as learning the tune?) Is there a very large number of processors available for executing programs in parallel, or only a small number (e.g. seven plus or minus two?) The former would allow arbitrarily complex hierarchically organised skills to be developed, subject possibly only to the constraint that a single global synchronising 'beat' is to be shared between them all. How deep can the parallel process hierarchies get? To what extent is horizontal communication across the hierarchies possible? What happens if the central processor and a low-level processor both attempt to run the same program? (Breathing seems to be an example where this might occur, since it is controlled intelligently in speaking, singing, etc. in addition to being an 'automatic' process.) Perhaps the running is always done by a lower-level processor, but sometimes under the control of the more intelligent program sythesiser? How are the primitive instructions routed from processors to still lower level processors, e.g. to muscles? If programs are physically copied into the lower level processors, then can processors be re-used during the process of development and debugging a skill? Is there some sort of garbage collection of processors? Similar questions arise about the space required for the alternative system where different processors access the same program stored in the same location. Can storage space for instructions be re-used? How are new processors and new storage space allocated? Do the different processors share limited resources of some kind, e.g. memory or 'fuel', or are they truly independent? Does this hierarchical parallel organisation of "motor" skills also play a role in other abilities, e.g. perception, language understanding, problem-solving? What are the implications of all this for our understanding of consciousness? Is it possible that many of the lower levels use computational resources of types which first evolved in much less intelligent organisms? How did the newer, more sophisticated mechanisms evolve?

This is a short list of questions I cannot now answer. I don't claim to have offered a theory. At best it is a research program which may produce explanatory theories one day.

References

J.A. Fodor, The Language of Thought, Harvester Press, 1975.

A. Sloman, The Computer Revolution in Philosophy: Philosophy Science and Models of Mind, Harvester Press and Humanities Press, 1978.

A. Sloman, 'Why robots will have emotions', submitted to IJCAI 1981.

Cognitive Style and the Learning of a First Computer Language

M.J. Coombs
Man-Computer Communication Research Group
Computer Laboratory
University of Liverpool
Liverpool, L69 3BX
U.K.

1. Individual Differences and Computing Skills

It is commonly asserted that the learning of computing skills is subject to marked individual differences (e.g. Weinberg, 1971; Gould, 1975; du Boulay and O'Shea, 1981). Such differences are of considerable importance for both practical and theoretical reasons. First, an analysis of successful and unsuccessful learning strategies should provide data on the nature of computing information and the skills required for its acquisition. This information would aid in the design of systems for guiding the rapidly increasing population of computer users who are not computer professionals. Secondly, the existence of clear individual differences should provide an ideal environment for exploring the relationships between learning strategy and cognitive performance.

Research was therefore undertaken with these two objectives in mind, but emphasising the first. Work was conducted in two stages:

the identification of the main parameters of learning style related to the learning of a first computer language;

the definition of interactions between style, strategy and knowledge representation.

2. Stage 1: Four Programming Studies

In order to characterize the relationship between learning style and the acquisition of programming skills and to provide early applicable results, research was conducted on science and engineering postgraduate students taking a standard introductory FORTRAN course (Coombs et al., 1981).

A research methodology was used in which subjects were presented with two tasks: a "target task" for the assessment of programming skill, and an "indicator task" for the assessment of learning style. The objective was to characterize performance on the target task, about which little was known, with reference to performance on the indicator task. It was thus important that the indicator task was well-founded and had some clearly defined relationship with the target task.

Two programming tests were devised for the target task, following the finding (Coombs, 1977) that recall of the units of procedural information may be independent of ability to recall the units in their correct order. The first test the Statement Test was devised to assess learning of the format and operation of individual FORTRAN structures and required the identification of appropriate errors in 3 short programs (21 in 56 lines of program). The second test the Logic Test was devised to assess ability to assemble individual structures into successful programs. Subjects were presented with 3 decks of cards, each comprising a complete program. The cards were given in random order and subjects were required to reconstruct the original program.

Selection of an indicator task proved difficult because of the poorly understood and unstable relationships between tests of cognitive style. A test was therefore used which had significant features in common with the skills assessed by the target tasks. A suitable classification of individual differences was found in the Operation Learning/Comprehension Learning distinction proposed by Pask (1975). The distinction was founded in a well-elaborated theory Conversation Theory and a test was available the Spy-Ring History Test.

Pask's categories are defined within two basic dimensions of human information processing: a) the management of data selected from the world (attention); b) the mental representation of that data (mental model building). The attention dimension draws the binary distinction between local and global features of a subject material. The representation dimension draws the distinction between representation of new information in terms of descriptions and in terms of procedures to be used for its generation from previously acquired information. Given interactions between the dimensions, Pask proposes two basic cognitive styles: operation learning, (procedure building using low-level, local information); comprehension learning (description building with attention to global features). Students biased towards operation learning are thus able to learn rules, methods and details but are unaware of how and why they fit together. Comprehension learners learn an overall picture of the subject matter but may not be able to perform the operations to use it. Comprehension of the information exists in the absence of rules or operational meaning and in ignorance of details required for the information to be used.

The procedural aspect of computing provides a conceptual link between cognitive style and the programming tests, it being our contention that learners with an operation bias would have a dual advantage of the Logic Test but no advantage on the Statement Test. With the Logic Test, operation learners would be expected to attend during the learning of a language structure to its internal logic. However, comprehension learners would only attempt to remember the global features of the structure, making no close distinction between the essential features and those arising from its illustrative context. A comprehension learner would be expected to accept the example as given and always work within the framework until he was presented with a different example. The operation learner would thus have an advantage when faced with the Logic Test, because he would have a clear idea of the essential features of the structure independent of context. Structures could therefore be readily assembled, without interference, to complete the problem program. The operation learner would also be expected to gain a second advantage by applying an operation strategy to the problem itself.

The Statement Test would not be expected to be so sensitive to learning style, all errors being contained within individual statements and not being related to the logic of the programs. They should therefore be identifiable from incidental learning of the surface features of the language. They would not 'look right' to anyone who had been exposed to FORTRAN, irrespective of his understanding of the workings of the language.

Four programming studies were conducted on different groups of students attending a standard FORTRAN course, the only addition being a request

3. Stage 2: Cognitive Style to Learner Strategy

The research reported above focussed on the role of cognitive style in learning. However, in order to apply the findings to designing educational and guidance systems for unsuccessful students, one must translate style into strategy. Protocol data therefore was collected during the first, third and fourth programming study on subjects' learning activities during the lectures and practicals. In the fourth study a computer based concept elicitation technique was used to record the changing understanding of concepts during the course. The protocol studies concentrated on subjects with an extreme operation learning (10) and extreme comprehension learning (8) bias.

The research has provided a description of two classes of strategic element: those concerned with mental processing and those concerned with learning activities. From these two classes it is possible to identify model strategies which are similar to competence models in linguistics. In the presence instance our model strategies also have sufficient empirical foundation to provide a basis both for planning further research and guiding instruction.

Learning to use any symbol system requires the student both to progressively identify acceptable structures and to apply them to some processing objective. Moreover, the competent use of a symbol system implies the ability to perceive structures at many different levels (Van Dijk, 1977). A symbol system, therefore, has both microstructure and macrostructure. Both levels should be born in mind when interpreting our findings. Green et al (1981) emphasises the importance of macrostructure for the writing and debugging of programs.

Operation learners appear to actively use both lecture and practical sessions. The former are used for recording facts, the learner ensuring that the lecturer's statements and examples are recorded accurately in full. Examples are also recorded carefully, the operation learner's priorities being a) to copy the example, b) to follow the lecturer's analysis of the logical relations between statements, c) to consider alternative methods of achieving the same programming objective. In b) and c) the operation learner works from "inside" the code, paying close attention to the logical relations between statements. By following these activities, the student leaves the lecture with an accurate and relatively uninterpreted record of the structures discussed and the examples given, and understanding the internal logic of most of the examples.

The major learning activity of operation learners takes place during practicals. Here they continue to build their knowledge of the functional relations between low-level language structures so leading to a knowledge of useful macrostructures. This again appears to be achieved from the "inside" by, for example, seeking various combinations of statements which will achieve the same processing objective. Operation learners thus build considerable knowledge of the computational possibilities of combinations of statements and rapidly develop a rich sense of macrostructure. The computer is itself the instructor, reference only being made to external information such as the principles behind the local implementation of FORTRAN or the design of the local operation system when prompted by a program failure which cannot otherwise be corrected. Through the practicals, the operation learner develops an understanding of the logic of program design and skills related to debugging and avoidance of errors.

Comprehension learners, however, adopt a very different distribution of activity. During lectures they attempt to achieve what the operation learner reserves for practical sessions. The comprehension learner performs significant "on line" processing and is rarely fully engaged in simply reporting facts and examples. He rarely takes verbatim notes, rather recording the abstracted or generalized results of his processing of the lecture material. Such processing is, of course, carried out without the aid of the computer. Significant structure is therefore determined by reference to external sources of information from related domains such as FORTRAN design principles or details concerning the local machine. Where relevant information is not available, the comprehension learner tends to create determining principles for himself rather than store the low-level information in unintegrated form. The comprehension learner thus leaves the lecture with represntations of language structures which have been determined to some extent by factors external to FORTRAN and which have never been validated by an actual computer run. Moreover, his knowledge of structure does not that students complete the Spy-Ring History Test early in the course and the programming tests at the end. A total sample of 42 subjects completed all tasks over a main study and four replications. The Spy-Ring Test yielded for each subject a score of operation, comprehension and incidental learning. The Logic Test was scored using a count of simple first-order transitional errors and the Statement Test was scored using the signal-detection theory measure of $P(\bar{A})$. Relationships were assessed using Pearson Correlation Coefficients.

The results for all these studies revealed one stable significant relationship (.05 level) between the exercise of operation learning and the Logic Test. Interpreting this result using Pask's (1975) theory and our own confirmatory factor-analysis of the Spy-Ring Test, it was concluded that the skill of assembling statements into programs was best acquired by subjects who:

- paid close attention to details;
- systematically abstracted the critical features of programming structures;
- represented structural relations in rule form.

Evidence of the relationship between cognitive style and the Statement Test was inconclusive, although contrary to our hypothesis there appeared to be a weak relationship with operation learning.

From results at this stage it was possible to make 4 assertions concerning the learning of FORTRAN:

1) One can define at least 2 different learning styles in a population of novice computer users.
2) Those exhibiting significant operation learning are more successful at assembling language structures into an effective algorithm.
3) The successful learning style is defined by close attention to detail and a preference for procedural representation.
4) Successful identification of individual language structures is independent of learning style.

develop beyond this point because he also fails to take advantage of practicals to conduct validation. The comprehension learner thus leaves a course with a descriptive knowledge of individual, low-level structures but no reliable, tested knowledge of macrostructure.

5. Some Final Remarks

The above accounts suggest that all students should be encouraged to use an operational strategy. However, Pask (1976) proposes that comprehension learners should be left with their natural style but with:

> accurate conceptual and supporting information;
> encouragement to solve problems in a structured environment.

It may be argued that the unique characteristics of FORTRAN and the specific nature of the Liverpool course make general proposals premature. Whilst accepting this possibility, we believe that the effects of operation and comprehension styles are sufficiently universal that similar differences will be found with other languages and computing situations. Although differences in detail may exist, we propose that whilst programming requires precise specification of code, and whilst languages fail to make macrostructure perceptually evident and are implemented on machines not easily modelled by the novice, similar effects will be found.

6. References

COOMBS, M.J. (1977). Modality and order recall in learning from television, with reference to teaching medical procedures. Ph.D. Thesis, University of Liverpool.

COOMBS, M.J., GIBSON, R. and ALTY, J.L. (1981). Acquiring a first computer language: a study of individual differences. In M.J. Coombs & J.L. Alty (eds), Computing Skills and the User Interface. London: Academic Press.

DU BOULAY, B. & O'SHEA, T. (1981). Teaching novices programming. In M.J. COOMBS & J.L. ALTY (eds), Computing Skills and the User Interface. London: Academic Press.

GOULD, J.D. (1975). Some psychological evidence on how people debug computer programs. International Journal of Man-Machine Studies, 7, 151-182.

GREEN, T.R.G., SIME, M.E. & FITTER, M.J. (1981). The art of notation. In M.J. COOMBS & J.L. ALTY (eds), Computing Skills and the User Interface. London: Academic Press.

PASK, G. (1975). Conversation, Cognition and Learning. Amsterdam: Elsevier.

PASK, G. (1976). Conversation Theory : Applications in Education and Epistemology. Amsterdam: Elsevier.

VAN DIJK, T.A. (1977). Semantic macrostructures and knowledge frames in discourse comprehension. In M.A. JUST & P.A. CARPENTER (eds), Cognitive Processes in Comprehension. New York: Wiley.

WEINBERG, G.M. (1971). The Psychology of Computer Programming. New York: Van Nostrand Reinhold.

THE CONCEPT OF E-MACHINE: ON BRAIN HARDWARE AND THE ALGORITHMS OF THINKING

Victor Eliashberg

Varian Associates, Palo Alto

The goal of this paper is to call in question the popular thesis that the problem of the algorithms performed by the brain (algorithms of thinking) has but little to do with the problem of brain hardware. The paper presents a simple example of a "brain-like" universal associative processor (referred to as E-machine) for which such a thesis would be obviously inadequate. More sophisticated examples of E-machines were studied in Eliashberg (1979).

1. A SIMPLE EXAMPLE OF E-MACHINE

Consider the "neural" network schematically shown in Fig. 1. The big circles with incoming and outgoing lines represent centers, elements which are assigned certain discrete coordinates in the network. The small circles denote couplings, elements whose place in the network is determined by a pair of coordinates of centers. A center may be viewed as a neuron with its dendrites and axon or a neural subsystem that may be treated as a "reduced neuron". A coupling may be interpreted as a synapse or a "reduced synapse", the white and the black circles corresponding to the excitatory and inhibitory synapses respectively. Using the terminology of Nauta and Feirtag (1979), the network of Fig. 1 may be characterized as a three-neuron nervous system (it has three neurons in a path between its input and output). Some simple animals have two- and even one-neuron nervous systems, so the network of Fig. 1 may be viewed as a morphological model corresponding to a rather advanced stage of the evolution of the brain. Accordingly, one may expect to get some interesting information processing characteristics in a neurobiologically reasonable functional model of this network. Therefore to reach the goal of this paper it seems sufficient to show why the traditional brain-hardware-independent approach to the problem of the algorithms of thinking would fail to adequately describe the psychological properties of an animal with such a simple nervous system. So much the more may this approach be inadequate in the case of human brain.

NOTATION.

$N_j(i)$ is the i-th center from the set N_j.

$S_{kj}(i,i')$ is the (i,i')-th coupling from the set S_{kj}.

ν is discrete time (the number of cycle).

$x(\cdot,\nu)$ is the input vector of the model. (The dot substituted for an index implies that the whole set of components corresponding to this index is assumed).

$G^x(\cdot,i)$ is the vector of gains of couplings $S_{21}(i,\cdot)$. This vector will be interpreted as the symbol stored in the i-th location of input long-term memory (ILTM) of the model.

$E(i,\nu)$ is a variable describing a hypothetical state of "residual excitation" (E-state) associated with center $N_2(i)$. Such a state might be interpreted as a certain phenomenological counterpart of the concentration of a chemical participating in a slow reversible reaction. Accordingly, in a more sophisticated model one may introduce several types of E-states. Such states will be viewed as the states of distributed "non-symbolic" short-term memory of the model.

$U_2(i,\nu)$ is the (postsynaptic) potential of center $N_2(i)$.

$J_2(i,\nu)$ is the output signal of $N_2(i)$.

$G^y(\cdot,i)$ is the vector of gains of couplings $S_{32}(\cdot,i)$ interpreted as the symbol stored in the i-th location of output long-term memory (OLTM) of the model. For the sake of simplicity we will assume that the state of long-term memory (ILTM and OLTM) is formed before the first moment of observation and doesn't change later, i.e. we will avoid the problem of learning.

$y(\cdot,\nu)$ is the output vector of the model.

In a rather general form a functional model of the network of Fig. 1 with the above input, state, and output variables may be described as the following machine.

$$y(\cdot,\nu) \leftarrow F_y(x(\cdot,\nu),E(\cdot,\nu),G^*(\cdot,\cdot))$$

$$E(\cdot,\nu+1) \leftarrow F_e(x(\cdot,\nu),E(\cdot,\nu),G^*(\cdot,\cdot)), \text{ where}$$

F_y is the otput procedure, F_e is the next E-state procedure, $G^*=G^x,G^y$. As it was mentioned above in this paper we are not concerned with the next G-state procedure and treat the variable $G^*(\cdot,\cdot)$ as a parameter.

For the goal of this paper it is sufficient to make rather simple assumptions about F_y and F_e. What is important for this goal is the presence of E-states rather than the details of their interaction with the input vector and the G-state and the details of their dynamics. With that in mind let us introduce the following "quasi-neural" description of F_y and F_e. (The reader with an appropriate background will be able to find many other descriptions of these procedures satisfying the requirements of this paper).

OUTPUT PROCEDURE, F_y :

The potential $U_2(i,\nu)$ is the following function of $x(\cdot,\nu),E(i,\nu)$ and $G^x(\cdot,i)$

$$(1) \qquad U_2(i,\nu) \quad S(i,\nu)\cdot[1 + \alpha\cdot E(i,\nu)], \text{ where}$$

$$(2) \qquad S(i,\nu) \quad \sum_{(j)} G^x(j,i)\cdot x(j,\nu) \ , \quad i=1,\ldots n_2$$

The layer of centers N_2 with lateral inhibitory couplings S_{22} performs the random equally probable choice of a center, $N_2(i_o)$, from the subset of centers with the maximum potential (Eliashberg, 1969, 1979). It is assumed that there is some noise.

$$(3) \qquad J_2(i,\nu) = \begin{cases} 1 & \text{if } i=i_o \\ 0 & \text{----} \end{cases}, \text{ where}$$

$$(4) \qquad i_o :\in M(\nu) = \{i / U_2(i,\ ) = \max_{(i)} U_2(i,\nu) > 0\}$$

We are using ALGOL-like notation, $:\in$, to denote the operator of random equally probable choice of an element from a set.

The output vector is determined as follows

$$(5) \qquad y(\cdot,\nu) \quad \sum_{(i)} G^y(\cdot,i)\cdot J_2(i,\nu)$$

NEXT E-STATE PROCEDURE , F_e :

The dynamics of E-state is described by the first order difference equation, the time constant, $\tau(i)$, of this equation depending on whether $E(i,\nu)$ increases, $\tau(i)=\tau^+$, or decreases, $\tau(i)=\tau^-$.

(6) $\tau(i)\cdot[E(i,\nu)-E(i,\nu)] = S(i,\nu)-E(i,\nu)$,

where

(7) $\tau(i) = \begin{cases} \tau^+ & \text{if } S(i,\nu) > E(i,\nu) \\ \tau^- & \text{if } S(i,\nu) \leq E(i,\nu) \end{cases}$

In what follows the system described by Exps (1) - (7) will be referred to as Model 1.

2. SOME GENERAL INFORMATION PROCESSING CHARACTERISTICS OF MODEL 1

2.1. Universality with respect to the Class of Combinatorial Machines

Let $x(\cdot,\nu), G^x(\cdot,i) \in X$, where X is a finite set of positive normalized vectors. Let $y(\cdot,\nu)$, $G^y(\cdot,\nu) \in Y$, where Y is a finite set of positive vectors. Let the number of locations of LTM be as big as required ($n_2 \to \infty$), so any desired software , $G^x(\cdot,\cdot), G^y(\cdot,\cdot)$, may be put into the LTM of Model 1 before the beginning of observation. Let $\alpha = 0$ (the mechanism of STM of Model 1 is "turned off").

It can be shown that in this case Model 1 can be programmed to simulate an arbitrary probabilistic combinatorial machine (with rational probabilities) with the input alphabet X and the output alphabet Y.

2.2. Universality with respect to the Class of Finite-State Machines

Let us split the input and output vectors of Model 1 each into two subvectors $x(\cdot,\nu) = (x_1(\cdot,\nu), x_2(\cdot,\nu))$, $y(\cdot,\nu)$ $(y_1(\cdot,\nu), y(\cdot,\nu))$, and let us introduce the delayed feedback $x_2(\cdot,\nu+1)$ $y_2(\cdot,\nu)$.

Let $x_1(\cdot,\nu) \in X$, $x_2(\cdot,\nu), y_2(\cdot,\nu) \in Q$, $y_1(\cdot,\nu) \in Y$, where X and Q are finite sets of positive normalized vectors, Y is a finite set of positive vectors. Let as before $\alpha=0$.

It can be verified that this modification of Model 1 can be programmed to simulate any probabilistic finite state machine (with rational probabilities) with input alphabet X, state set Q, and output alphabet Y.

2.3. E-States as the Mechanism of Mental Set

Let $\alpha>0$ (the mechanism of STM of Model 1 is "turned on"). Let us assume at first that the time constants are very big ($\tau^+ \to \infty$, $\tau \to \infty$) so the E-state of Model 1 does not change considerably during the interval of observation. Let other conditions be as in section 2.1.

Suppose the LTM of Model 1 contains all possible input/output pairs (associations) from $X \times Y$, i.e., for all $(a,b) \in X \times Y$ there exists $i \in \{1,..n_2\}$ such that $G^x(\cdot,i)=a$ and $G^y(\cdot,i)=b$. It can be shown that for any (deterministic) combinatorial machine with the input alphabet X and the output alphabet Y there exists an initial E-state, $E(\cdot,0)$, of Model 1 such that this model in this state simulates the above machine.

Thus being observed as a black box, Model 1 with a fixed state of its LTM may appear to an experimenter as any of m^l deterministic combinatorial machines ($l=|X|$, $m=|Y|$) depending on the "state of mind" ("mental state"), $E(\cdot,\nu)$, of this model. This result can be naturally extended to the class of (deterministic) finite-state machines in the case of the Model 1 with the feedback of section 2.1.

Let us remove the condition of infinite time constants but still assume that these constants are rather big. In this case the E-state of Model 1 will change slowly ("adiabatically") so this model will gradually change into different combinatorial machines. Thus the "non-symbolic" expressions (6),(7),determined at a hardware level, controll the "personality" of Model 1 as a symbolic machine.

2.4. Why Model 1 Jeopardizes a Brain-Hardware-Independent Approach to the Problem of the Algorithms of Thinking

Imagine a researcher trying to develop a theory of the behavior of Model 1 per se (Model 1 treated as a black box) without being concerned with the hardware phenomena in this model. Let us assume that the researcher is used to work with a hardware-independent higher level language, say LISP, and deal with traditional symbol manipulation concepts. It is very likely that Model 1 would fool such a researcher by pretending to behave as a "classical" symbolic machine. The researcher with the above mentioned methodological mental set and the knowledge base associated with the "classical" symbolic information processing paradigm would have little chance to think about the "non-symbolic" E-states of Model 1, much less to find the hardware expressions (6),(7) describing the transformations of these states. It is not difficult to imagine how a sophisticated chemistry of the brain (see, e.g., Iversen, 1979, Kandel, 1979) may lead to non-symbolic brain hardware expressions much more complex than Exps (6),(7). Thus the above mentioned researcher would hardly have a better chance to come up with an adequate theory of human mental states than he (she) does in the case of Model 1.

3. METHODOLOGICAL REMARKS

3.1. On the Whole Brain and the Parts of its Behavior

Let (A,s_o) be a hypothetical machine corresponding to the human brain, A, in its initial (roughly newborn) state, s_o. After several years of learning (A,s_o) is changed into an intelligent system (A,s_n). Model 1 and the more sophisticated examples of E-machines studied in the Eliashberg (1979) manuscript give reasons to believe in the following methodological thesis.

There may exist a relatively simple description of (A,s_o) in terms of a machine with the state set similar to that of the brain. There may be a good possibility to find this description by trying to answer in a single context a large enough set of specially selected basic neurobiological and psychological questions. At the same time it may be hopeless a strategy to try to find adequate descriptions of some nontrivial "parts" of the behavior of (A,s_n) without looking for a description of the "whole" system (A,s_o).

3.2. On the General Relation between the Theory of the Brain and the Information Processing Psychology

To clearly understand the implied methodological meaning of the above mentioned thesis it is useful to compare the general relation between (A,s_o) and the information processing psychology with the general relation between the basic equations of a traditional physical theory and this theory. As an example of the latter relation let us take the Maxwell equations and the classical electrodynamics. (Don't take this metaphor too literally!).

The relatively simple Maxwell equations allow one to describe all variety of arbitrary complex classical electromagnetic phenomena as a result of interaction of these equations with the correspon-

ding variety of "external worlds" represented in the form of various boundary conditions, media, and sources. To find the Maxwell equations did not mean to solve all the problems of classical electrodynamics, the latter having developed (and continuing to do so) a number of specific problem-oriented models and concepts. It would hardly be possible, however, to adequately formulate all these specific problems, let alone solve them, without the Maxwell equations.

3.3. On Skipping "Simple" Problems in order to Solve Complex Ones Faster

For the goal of this paper it is especially important to emphasize that the Maxwell equations were found in an attempt to adequately formalize and extrapolate some "simple" basic knowledge about electromagnetic phenomena (Faradey law,etc.). This formalization and extrapolation created a powerful mathematical tool allowing to adequately approach complex problems of classical electrodynamics.

Imagine a physicist trying to develop a computer program for simulating the behavior of electromagnetic field in a complex microwave device without being concerned with the equations describing the fundamental properties of this field. A researcher trying to skip "simple" basic neurobiological and psychological questions, in order to solve complex problems of human information processing faster, may well be in a situation similar to that of such a physicist.

ACKNOWLEDGEMENTS

I want to express my gratitude to Prof. H. Landahl, Prof. J. McCarthy, Prof. H. Martinez, Prof. I. Sobel, Prof. L. Stark, and Prof. L. Zadeh for useful comments.

REFERENCES

Eliashberg, V.M. On a class of learning machines. Leningrad, USSR, Proc. of VNIIB, 54, 1969 (in Russian).

Eliashberg, V. The Concept of E-machine and the Problem of Context-Dependent Behavior. Palo Alto, Ca., 1979, Copyright 1980, by V. Eliashberg, TXu 40-320, US Copyright Off.

Iversen, L.L. The chemistry of the brain. Sc. American, Sept. 1979.

Kandel, E.R. Small systems of neurons. Sc. American, Sept. 1979.

Nauta, W. & Feirtag, M. The organization of the brain. Sc. American, Sept. 1979.

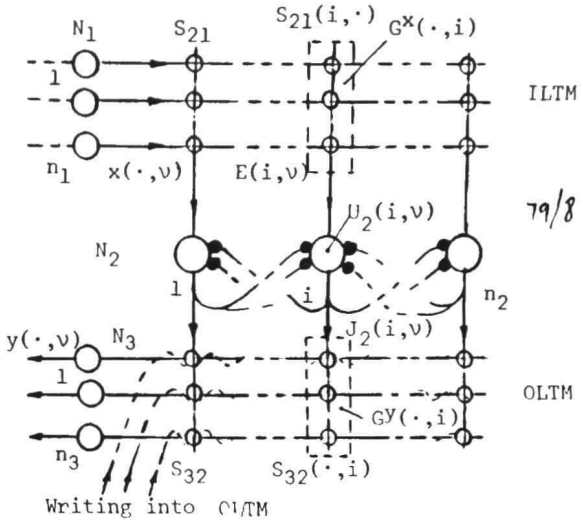


Fig. 1

Invariance Hierarchies in Metaphor Interpretation

Jaime G. Carbonell
Carnegie-Mellon University
Pittsburgh, PA 15213

Abstract

Interpreting metaphors is an integral and inescapable process in human understanding of natural language. The current investigation analyzes analogical mappings underlying metaphors and their implications for inference and memory organization. Regularities have been observed indicating that certain types of conceptual relations are much more apt to remain invariant in analogical mappings than other relations, resulting in an induced invariance hierarchy. The central thesis is that human inference processes are governed by the same analogical mappings manifest as metaphors in language.[1]

1. Introduction

Metaphor is a pervasive phenomenon in almost all human written and spoken language [7, 9, 6]. Recently, I proposed a model of metaphor comprehension based on the identification of general metaphor mappings and subsequent recognition of metaphors as instances of previously-encountered generalized mappings [3]. This method was meant to be a computationally-effective means of interpreting "common" or "mundane" metaphors. As such, it did not address issues of how the underlying analogical mappings structure inference processes; nor did it consider the more difficult task of understanding truly novel metaphors. Here, I investigate the inference processes underlying novel-metaphor comprehension and their implications for memory organization.

Recognition and initial interpretation of metaphors in written or spoken language represents only the tip of the proverbial iceberg. Metaphors in narratives, dialogs and informative text are the observable linguistic manifestations of a central, underlying cognitive process. My thesis is that *cognition is dominated by analogical reasoning* [4], in contrast to more rigid reasoning models based on "sounder" principles of formal logic (e.g., deduction, resolution, etc.). Metaphor is the reflection, on the language medium, of analogical thought processes; as such it provides essential clues of the inner functioning of human inference processes. This paper discusses the utility of metaphor as a tool to investigate various cognitive processes.

2. Two Metaphors are Better Than One

Consider the metaphor ***inflation is war***, as discussed by Lakoff. Newscasters talk of "*fighting* inflation", "workers *taking a beating* from inflation", "Carter *loosing another round to inflation*", "inflation *overrunning* our economy", "savings being *attacked* by inflation", etc. Of what possible use is this metaphor to the reader (or the writer)? Inflation is an economic phenomenon whose causes, implications, and methods of control are not understood by the public at large[2]. (Indeed, some would say that inflation is poorly understood by politicians, economists and business men alike.) Therefore, the metaphor helps to enrich the knowledge brought to bear in the comprehension process by transfering corresponding appropriate information from the more familiar adversary-conflict situation. The necessity to enrich and elaborate upon concepts in the understanding process has been amply demonstrated by Bransford and Johnson [2], Anderson [1] and others. The only feasible way to bring knowledge to bear on an ill-understood domain is to construct a metaphor suggesting a useful transfer mapping of factual information and inference rules from a better-understood domain.

Given the general notion that metaphors transfer knowledge from well-understood to more ill-structured domains, three questions arise:

1. How can the transfered knowledge be used? (i.e., what does the metaphor provide that a literal description may lack?)
2. How does the transfer of knowledge actually take place? (i.e., what would constitute a computationally-viable mechanism?)
3. What implications does the utility and pervasiveness of metaphor have on cognitive processes such as memory organization, inference, and learning?

We consider each question in turn.

3. Why Metaphor?

The knowledge transfered from a richer domain to a more impoverished one via an analogical mapping, triggered by the use of linguistic metaphor, plays a key role in inference processes. Let us return to the ***Inflation is War*** metaphor. If a newspaper article opens with this metaphor, the entire text can be organized around it. Equating inflation with a personified "enemy" enables one to draw upon the knowledge organized under "adversary conflict" to suggest courses of action (in terms of the organizing metaphor)[3]. For instance, we can now understand that in order to "*vanquish* inflation", we clearly must: "formulate a *battle-plan*", "*marshal our forces*", "*take the initiative*", "*go on the offensive*", and "make a *determined attack* on inflation" in order to "stamp it out of our society" and remain on the alert for "future *bouts* with inflation. In short, we must "*whip inflation now* (WIN)" as President Ford said when he "*launched his campaign against inflation*" When the metaphor has been drawn, it is reasonably easy to formulate subgoals based on the better-understood source domain. This is the first step in planning purposive behavior.

The inferences one can draw on how to deal with inflation are all structured by the initial metaphor. Different metaphors will yield markedly different sets of inferences. In order to illustrate that there is nothing inherently special about the ***inflation is war*** metaphor, consider another metaphor used to discuss inflation, encountered frequently in the Spanish press, but easily understood once stated: ***Inflation is a disease***. Here, the economy is the patient, inflation is the infecting organism that must be driven out of the patient with the help of the physician (the economist who sets national economic policy).[4] Hence, one can "*take the pulse* of the economy", "*diagnose* the cause (always placing the blame on external forces - just as disease organisms are an external cause of illness)", "*prescribe treatment*", "put the economy on a *lean diet*", "make the *medicine palatable* to the poor", "wait for private enterprise to *recuperate*", "perform *radical surgery*, cutting *swollen* budgets", "*treat the symptoms* while the inflation continues to ravage", and "*relieve the pain* by subsidizing the price of necessities"

---

[1] This research was sponsored in part by the Office of Naval Research (ONR) under grant number N00014-79-C-0661, and in part by the Defense Advanced Research Projects Agency (DOD), Order No. 3597, monitored by the Air Force Avionics Laboratory under Contract F33615-78-C-1551. The views and conclusions contained in this document are those of the author, and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

[2] The popular adage that inflation is "Too much money *chasing* too few goods" is itself a metaphor, one that suggests different corrective action.

[3] This realization is due to Lakoff.

[4] One can imagine this author's confusion upon reading in a Spanish newspaper about the "national malady", the proposal to "inoculate workers" by cost-of-living adjustments, and a "prescription for the national health". What sort of epidemic was on the loose? However, once the metaphor was understood, the text made perfect sense. Since this metaphor ***is the way in which inflation is always discussed***, there appeared to be no need to introduce it explicitly. Moreover, no one would admit that inflation was not a disease, as the metaphor so deeply permeated discussions of inflation that metaphorical terms were not recognized as such. In a conversation with local person, the following statement was made in reaction to my statement suggesting inflation was being discussed in terms of a medical metaphor: "*Of course*, our economy is sick and must be cured, literally! I mean just what I said." This episode should help us step outside our own metaphor and realize that no one can literally *battle* inflation, but that the metaphor is so ingrained in our thinking that we can draw inferences and make statements easily *only if* we rely on the accepted metaphor to structure our reasoning processes.

The moral that can be drawn from these two examples is twofold:

1. **Inference and planning are directly structured by the analogical mapping underlying a dominating metaphor.** The first inflation metaphor suggested *tactics* against inflation, whereas the second suggested *cures* for inflation. Therefore, the inference mechanism consists of mapping corresponding solutions from the source to the target domains. Hence, the metaphor must equate two problems, one of which is better understood and therefore suggests inferences and plans presumed applicable in solving the second problem. (See [4] for a detailed discussion of analogical problem solving.)
2. **Solutions to problems generated by metaphors are ONLY useful as heuristic problem-solving advice.** No **detailed** solutions in ***war*** or ***curing disease*** can be applied directly to inflation. How would one "shoot bullets at inflation" or "get inflation to sign the Geneva convention"? Similarly, one cannot "intern the economy in a hospital" or "give it an intravenous penicillin injection" Clearly, the underlying analogy does not extend to the object level. However, the *planning* level provides useful information. Therefore the **intentions of the actors** are preserved in the mapping, as is the **causal structure** of the events, but the **instantiations of** the events themselves are lost in the analogy. This observation accords with Winston's analogy mappings based on preservation of causal structure [12] and Gentner's discussion of analogy in scientific theory [5].

4. How Metaphors Structure Inference Processes

As we discussed in the preceding section, metaphors can establish an expectation setting for comprehending large portions of text. This expectation setting is generated by transfering inferences from the source to the target domain, including the goals and plans that actors in the target domain are expected to pursue. (It is important to realize that the goals of "defeating" or "curing" inflation come from the respective metaphors -- not from the concept of inflation itself. Therefore a language understanding system must tap the metaphor to comprehend exactly what problems are caused by inflation, and what their respective solution strategies ought to be.)

Let us define *creative metaphors* to be the linguistic realizations of large scale analogical mappings that can structure entire planning episodes. Creative metaphors include the two inflation examples above, Gentner's scientific-theory metaphors [5], and each of the roughly 50 generalized metaphors discussed in [3]. Non-creative metaphors are frozen metaphor instances with fixed meanings, or figures of speech (such as "kick the bucket") whose metaphorical roots can only be traced through their etymology. Non-creative metaphors do not map inferences, as their source domain has been lost in their history, and therefore is not available to the understander. In short, if a metaphor enables one to bring knowledge to bear from an existing domain to a new, less understood domain, we define it as a creative metaphor. The discussion below centers on the process that brings knowledge to bear in understanding new domains.

In order extract information from an existing domain to a new domain via a metaphor, it is crucial to know what aspects of the existing domain should remain invariant in the mapping, which should be transformed, and which should be ignored. As we saw in the previous section, objects are seldom, if ever, preserved in a metaphorical mapping, whereas planning structures are mapped invariant to the new domain in fact it is precisely because planning structure and inferences can be preserved by analogical mappings that metaphors are powerful means of helping an understander formulate reasonable behavior in uncharted domains.

An analysis of some two hundred creative metaphors yields the following empirical observation. There is a well-defined invariance hierarchy among the aspects of a situation that are mapped by a metaphor. This perceived regularity is remarkably consistent across metaphors in different domains. In fact, metaphors that are rated as "bad metaphors" often violate the invariance hierarchy presented below.[5] Hence, a plausible hypothesis is in that people *expect* certain aspects of the source domain to remain invariant and other aspects to be coerced into corresponding entities in the new domain. These expectations can focus the search for metaphor interpretations. The regularities observed over a large number of metaphors are summarized by the normative invariance hierarchy presented below. The conceptual relations in the hierarchy are listed in decreasing order of expected invariance:

- *A goal-expectation setting for the animate actors involved (if any).* Goals, if present in the source domain, are almost always mapped invariant into corresponding entities in the target domain. If the source domain contains animate actors and the target does not, then the goals of the actors will be attributed to the corresponding personified entities in the target domain. E.g., inflation becomes an anthropomorphized malevolent agent in ***inflation is war***, therefore the *the goals* or a nation at war are mapped invariant in that inflation must be *fought and defeated*.
- *Planning and counterplanning strategies among competing or cooperating actors.* -- These strategies, almost always preserved intact by an analogy, provide a priority ordering among the goals and suggests possible means for pursuing each goal. Often, the most useful aspect of a metaphor is to enable purposive planning in what previously was too ill-structured a domain.
- *Causal Structures.* -- When the causal structure of the source domain is explicit, it will typically be preserved by the mapping. E.g., medicine cures disease; therefore economic measures will "cure" inflation. In Reddy's conduit metaphor for how people talk of language [9], causal structure abounds. E.g., a blocked conduit prevents physical transfer; therefore press censorship will also block dissemination of ideas.
- *Functional Attributes.* -- The function to which an object in the source domain is typically applied will often be coerced onto an analogous function for the corresponding object in the target domain.
- *Temporal Orderings.* Normative planning sequences in the source domain map into potentially applicable planning sequences (instantiated differently) in the target domain, often preserving temporal relations.
- *Natural tendencies.* -- In the celebrated analogy between electric circuits and and hydraulic systems (used to explain Ohm's Law), water "tends" to go down hill, therefore electricity "tends" to go towards the voltage "*drop*" Moreover, thin pipes resist the flow of water, therefore thin wires "resist" the "*flow*" of electricity.[6]
- *Social roles.* -- Social relations are sometimes preserved and sometimes not. In a battle there are generals and foot-soldiers; therefore, the war against inflation must be fought by many wage-earners (soldiers) under the direction of economic planners (generals). Since doctors cure the disease directly, the economic planner must shoulder the entire burden, and wage-earners (patients) are essentially powerless with respect to taking an active role in the cure. Both mappings preserve the inferences associated with the social roles in the source domain. However, the more specific roles of "spy" and "submarine commander", are not preserved by the ***inflation is war*** metaphor.
- *Structural relations.* -- Occasionally structural relations remain invariant in an analogy, but often they are transformed or suppressed. For instance, in the Rutherford solar-system model of the atom, physical relations between the electrons (planets) and nucleus (sun) are remain invariant. (In both case there is an orbit relation as a function of an inverse-square centripetal force). However, Saying "John is at the head of his class" does not preserve the physical structure normally found between a body and a head (the latter being connected to and nourished by the former).
- *Descriptive Properties.* -- These are the least likely properties to be preserved in a metaphor. Wires and pipes are both long and narrow (in the hydraulics metaphor) However, Generals

---

[5] This observation is based on data collected by Lakoff, Gentner, Ortony and others

[6] Electricity is actually not a flow of electrons, but we always think of it that way because the hydraulics metaphor pervades our discussions of electrical phenomena

are military men, economic planners are usually academics. The sun is yellow-orange, very large and has sunspots, but as Gentner points out, none of these monadic descriptors apply to the nucleus of the Rutherford atom.

- *Object identity.* Objects in the source domain are almost never mapped onto objects of identical type in the target domain. Therefore, there are no tanks, bullets, M16's, attack submarines, uniforms, or field hospitals in the battle against inflation.

5. Analogical Mappings in Problem Solving

Our discussion suggests that metaphors are a useful means of indexing mappings between goals, planning structures, causal connections, tendencies, relations, and descriptions (in decreasing order of invariance and significance). Not all components are present in every metaphor. The preferred-invariance ordering helps us understand how metaphors may be used to facilitate reasoning processes in new domains, namely:[7]

1. Establish the invariant components of the mapping
2. Establish initial correspondences among the entities in the source and target domains. (This is a very partial correspondence -- only entities that are referenced by an invariant component in the explicit mapping can be directly related.)
3. Goal-correspondence identifies the problems that must be solved in the target domain. [What should one do about inflation? The ***disease*** metaphor metaphor states that it should be eradicated. The ***war*** metaphor suggests subjugating it. A comparison of inflation with an overindulging *gourmand* would yield the goal of trimming it down and controling its scope, but not eliminating it.] Therefore, metaphor actually determines the goals that one ought to pursue in the target domain. Without knowledge of goals little purposive action can take place (i.e., problem solving becomes meaningless, as there are no goal states in the problem space.)
4. Planning strategies invariant under a metaphorical mapping transfer operators from the source to the target domain, hence establishing a problem space [8] and suggesting potentially troublesome interactions among operator preconditions. The ***inflation is a disease*** metaphor suggests that since administering medicine is a useful operator in the medical domain, a correlate operator ought to be useful in the economic domain. Moreover, medicine is usually an unpleasant experience, therefore the inference is made that its economic correlate would be unpleasant as well. Hence, we speak of giving the economy a strong *dose* of anti-inflationary monetary restraints, and making the policy *palatable* to workers.
5. Causal connections classify operators in the target domain by the differences they reduce (analogized from the source domain) "The pressure of the water is determined by the product of the rate of flow and the cross-section of the pipe" suggests that in order to know the voltage, one can measure the current and resistance. Therefore a way of reducing the KNOW-V goal is to apply the multiply(I, R) operator, reducing the KNOW-V goal to the subgoals KNOW-R and KNOW-I.
6. Natural tendencies, social roles and structural relations provide information about the applicability conditions of operators [E.g., who can administer medicine (decide economic policy)?], and provide heuristic guidance to planning processes in the new domain. [E.g., Wars are costly and people must make personal sacrifices; therefore in battling inflation the cost should be taken into account, and the planner should be ware of potential problems caused by those who are unwillingly called upon to make the sacrifices.]
7. Temporal-progressions suggest macro-operators (typically useful sequences of operators) In treating a disease we first must identify the cause, then prescribe medicine then wait patiently for it to take effect. In war we marshal our forces (no searching for a suitable the enemy is needed, as the enemy is known at the start of hostilities), then attack (no waiting for the attack to take effect is necessary). Therefore we see two very different general plans suggested by the two metaphors. However, recall that the metaphors shared the same general goal. It is typically the case that most metaphors used to explain a particular ill-defined situation will share common goals and diverge increasingly as one traverses down the invariance hierarchy.

Reiterating the central theme of this section: metaphors provide a problem space, including a goal state, operators indexed by differences they may reduce, and normative plans that may prove useful. In essence, they make problem solving possible in what may previously have been too ill-structured a situation to make any progress. Metaphors do ***not***, however, provide any ***canned*** solutions applicable directly to new problems such would be an unreasonable expectation.

6. Exploiting the Invariance Hierarchy

The invariance hierarchy provides a first-pass solution to an apparently simple phenomenon that had perplexed some investigators, including this writer. When we hear that "John is a fox" we interpret it to mean "John is sly", not "John has pointed ears and a bushy tail." Similarly, we interpret "John is a pig" as a remark on his personal habits or his obesity, rather than a statement that John lives in a farm and has a curly tail. A partial answer to this problem lies in knowing the most salient feature of the animal to whom a person is compared. However, a more complete answer is provided by exploiting the invariance hierarchy in the following manner: Consider the animal (source domain) and scan down the hierarchy stopping at the first entry for which we have a commonly known fact. For foxes we stop at planning/counterplanning -- folk wisdom tells us that foxes are very adept at devious counterplanning behavior. Hence, we never reach the physical descriptors of a fox. For pigs we may stop at either "natural tendency" (if we believe that pigs tend to get fat) or at "social role" -- folk wisdom asserts that pigs play a distinct role in the animal kingdom as the least hygienical of all animals. If we heard "John is a Giraffe", we find no common knowledge anywhere in the hierarchy until we reach physical attributes. Here we pick the most salient ones (e.g., height and/or length of neck), to understand the metaphor. The key to the process is that comparisons along the higher-invariance entries in the hierarchy are preferred. Once a high invariant property is found, no lower ones are considered. This is crucial to understand "John is an elephant" as a remark on the length of his memory (or his capacity for work), not the length of his nose (trunk), although the latter is perhaps the single most salient feature of elephants. Physical descriptors, however, are ranked low in the hierarchy.

7. Implications for Memory Organization

We have outlined how reasoning based on metaphors may proceed. Now, consider another aspect of metaphorical reasoning: *How are metaphors formed in the first place?* Given the ubiquity of metaphor, it becomes strikingly apparent that humans generate metaphors as readily as they understand them, occasionally unconscious of the fact that they are creating (or more often instantiating) metaphors. The question that must be posed is more specific: *What memory organization could enable, facilitate and encourage the continuous creation of metaphors?*

If we assume that the invariance hierarchy is roughly correct, it provides a best-first criterion for searching a content-addressed episodic memory, organized along the general lines of Schank's MOPS [11, 10]. In investigating reminding and inference phenomena, Schank asserts that detecting similarities at every level of abstraction is the key to human memory organization. Accepting this notion requires one to have a means of computing similarities among large numbers of potentially-relevant episodic traces, both for memory access and update. The hierarchy above suggests that goal similarities are crucial, planning-level similarities are almost as important, and similarities across other dimensions are of progressively lower importance. Hence, if memory were organized according to the computational criteria required for metaphor comprehension, it follows that a

[7] Here I adopt Newell and Simon's Means-Ends Analysis framework for problem solving [8]. The reader is referred to [4] for a more detailed discussion of analogical problem solving.

hierarchical structure would result, where the categories formed are largely determined by groupings along the entries in the invariance hierarchy, the more invariant entries corresponding to more global organizing categories. The actual content of the hierarchical memory is determined primarily by the idiosyncratic experience of the individual. Therefore, memory searchers for "good metaphors" (those preserving high-invariance properties) require less work (either to generate or comprehend) and may prove more rewarding for the understander as they index relevant memory more readily.

Metaphor is a linguistic realization of an inference phenomenon. As such, it should reflect underlying memory structure, as well as suggest the types of inferences people can perform most readily. If we ask *why* creative metaphors are used, the most logical answer appears to be that the writer is trying to induce the reader to perform the necessary inferences required to comprehend the new material. Metaphor serves as a vehicle to suggest a fruitful domain from which the relevant inferences can be mapped onto the new domain. Hence, when Senator Joe McCarthy refered to Communism as a "dreaded plague", he was inducing, in the minds of his listeners, the inference that communism must be actively "eradicated" or it will spread. The metaphor is effective only because the appropriate inference structure was already in existence in the source domain, and McCarthy knew this at the time he created the metaphor.

An interesting avenue of future research is automating metaphor generation. If the model discussed here is essentially correct, metaphor generation requires that the writer have a model of the knowledge state (including goals, strategies, beliefs, etc.) of his reader, as well as an integrated episodic memory where the computed similarity metric incorporates the invariance hierarchy. (I.e., two domains are considered similar if the same types of problems and inference processes are present in planning effective behavior in both domains.)

In order to clear possible misconceptions, I emphasize that no distinct, localized, "conscious" existence for the invariance hierarchy is postulated as part of a human memory model. My hypothesis is that *the regularities manifested in the hierarchy are epiphenomenal reflections of human memory organization and inference mechanisms.* As such, the invariance hierarchy summarizes a phenomenon that must be explained by comprehensive memory-organization models, and hence it ought to be taken into account in the model-formulation process.

References

1. Anderson, J. R., *Language, Memory and Thought,* New Jersey: Erlbaum, 1976.

2. Bransford, J. D. and Johnson, M. K., "Considerations of Some Problems of Comprehension," in *Visual Information Processing,* W. G. Chase, ed., New York: Academic Press, 1973.

3. Carbonell, J. G., "Metaphor - A Key to Extensible Semantic Analysis," *Proceedings of the 18th Meeting of the Association for Computational Linguistics,* 1980 .

4. Carbonell, J. G., "Learning by Analogy: Skill Acquisition in Reactive Environments," in *Machine Learning,* R. S. Michalski, J. G. Carbonell and T. M. Mitchell, eds., Palo Alto, CA: Tioga Pub. Co., 1981.

5. Gentner, D., "The Structure of Analogical Models in Science," Tech. report 4451, Bolt Beranek and Newman, 1980.

6. Hobbs, J. R., "Metaphor, Metaphor Schemata, and Selective Inference," Tech. report 204, SRI International, 1979.

7. Lakoff, G. and Johnson, M., *Metaphors We Live By,* Chicago University Press, 1980.

8. Newell, A. and Simon, H. A., *Human Problem Solving,* New Jersey: Prentice-Hall, 1972.

9. Reddy, M., "The Conduit Metaphor," in *Metaphor and Thought,* A. Ortony, ed., Cambridge University Press, 1979.

10. Schank, R. C., "Reminding and Memory Organization: An Introduction to MOPS," Tech. report 170, Yale University Comp. Sci. Dept., 1979.

11. Schank, R. C., "Language and Memory," *Cognitive Science,* Vol. 4, No. 3, 1980 , pp. 243-284.

12. Winston, P. H., "Learning and Reasoning by Analogy," *CACM,* Vol. 23, No. 12, 1979 , pp. 689-703.

DUAL-SYSTEM PROCESSING IN LANGUAGE COMPREHENSION

Jon M. Slack
Open University, England

1 Imagery and Language Understanding

Natural language parsers map linguistic input strings into symbolic, relational structures using syntactic knowledge, semantic knowledge, or both. However, no parser maps such inputs into image codes which implies that people building language understanding systems do not regard image generation and processing as a necessary part of language understanding. Within psychology, on the other hand, some researchers have advocated a strong relationship between image processing and language understanding for the past decade [1]. The status of this view is by no means unequivocal amongst psychologists, and the debate about imagery has littered the pages of many an academic journal. To facilitate your reading of this paper it is worth stating that the work reported here is driven by the belief that image processing is an essential component of language understanding.

2 Using Image Codes in Language Understanding

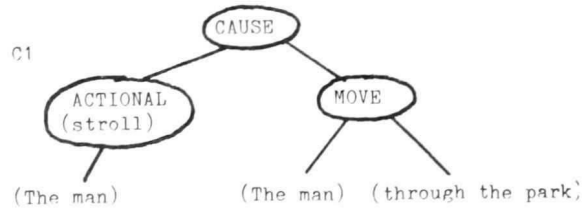
If you ask people to outline the difference in meaning between the words strolling and sauntering they find the task relatively difficult. If you ask them how they arrived at their description, the majority of people state that they formed images of a person strolling, and sauntering, then compared these images generating the best verbal description of the differences between the two they could. This sort of protocol data suggests that it is necessary to process image codes in order to distinguish between the meanings of the two words. That is, at least part of the meaning of the words is represented in some form of image code. If this is the case, then for a natural language understanding system to distinguish between sentences S1 and S2 it needs to make recourse to the image-coded

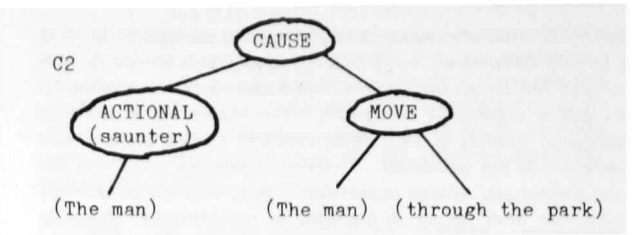
S1 The man strolled through the park.

S2 The man sauntered through the park.

components of the two verbs. This is not saying that an understanding system needs to map the sentences directly into image codes in isolation of other semantic structures. Rather, image processing is an essential component of the total processing involved in building cognitive structures which represent the differentiated meanings of the two sentences.

Most existing language understanding systems parse linguistic inputs into some form of propositional network which represents the conceptual relations corresponding to the meaning of the input [2],[3],[4]. These systems never use non-propositional codes, that is, language understanding is totally contained within a propositional system. Using the work of Norman and Rumelhart [5] as an example system, sentences S1 and S2 would be parsed into the conceptual structures C1 and C2, respectively.



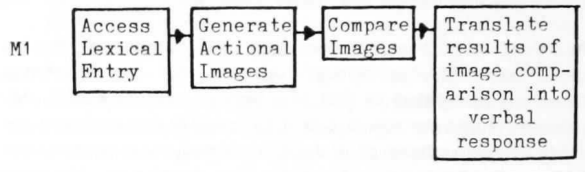


Within this system the two sentences are differentiated in terms of the actional components which represent the physical actions which implement the movement component implicit in the meanings of the verbs. These actional components are associated with the image codes used in distinguishing the verbs.

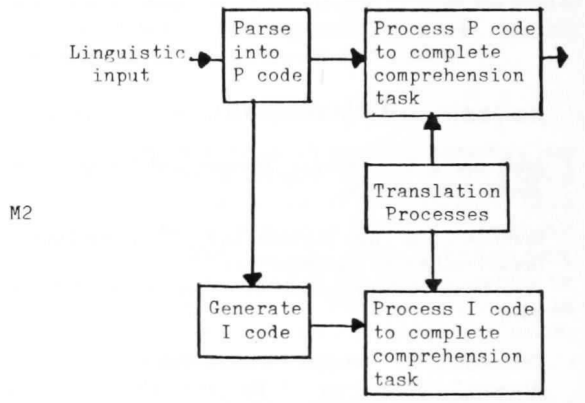
Kosslyn and his fellow researchers [6] have developed a detailed theory of image generation and processing which is backed-up with a working computer simulation. He proposes that numerous cognitive tasks, in partiuclar size-comparison and sentence-verification tasks, involve the processing of both propositional and image codes. The two types of codes and associated processes constitute independent, but connected, processing systems which run in parallel. Kosslyn's theory provides a good foundation on which to build a more adequate model of language understanding incorporating image processing in addition to existing ideas of propositional processing.

3 Model of Language Understanding

The protocol data described in section 2 suggests that the processing underlying the meaning-differentiation task can be accounted for by model M1.



In line with the ideas embodied within Kosslyn's model, a dual-system model of language understanding would have a structure similar to model M2.



In M2 the Propositional code system (P-system) and Image code system (I-system) are not strictly independent in that linguistic inputs are not mapped directly into images. Rather, the I-code is generated as a product of the parsing mechanism which maps the input into a P-code. However, once the two codes are created they are processed separately, but knowledge can be passed from one system to the other by means of the translation processes. Model M1 is accommodated within model M2 by the I-system and translation processes.

But what arguments are there for advocating the addition of an image processing system to existing language understanding models?

Argument 1: the I-system has either sole access to, or more direct access to, knowledge which is

necessary for efficient language comprehension. This knowledge maybe represented only as I-codes. On the other hand, it could be coded within both the I-system and P-system, but more accessible as images.
Argument 2: the I-system is more efficient at processing large amounts of knowledge within specific task domains. For example, I-codes provide a more natural medium for processing knowledge of spatial relations [7]. Further, Waltz [8] has argued that event simulation is an important aspect of language understanding, allowing the comprehension system to make inferences about scenes and judge the plausibility of inputs during parsing. The I-system is necessary for efficient processing of such event simulations.

4 Support for dual processing in comprehension

The imagery value of sentences has been shown to exert a profound influence on comprehension and memory tests [9],[1]. Thorndyke [10] has shown that subjects' imagery ratings for sentences vary inversely with comprehension RTs. Kosslyn's research project team have shown that imagery is used in sentence-verification tasks and for accessing knowledge of physical dimensions of objects. Further experimental evidence in support of Argument 1 is described below.

Experiment 1 Subjects were shown simple subject-verb-object sentences and asked either to read them for understanding (condition 1) or to generate an image corresponding to the meaning of each sentence (condition 2). In condition 1, the comprehension RTs were measured. In condition 2, the image-generation time were measured. Following the presentation of the sentences, subjects in both conditions were given a recognition test. For each target sentence in the test there were two distractor sentences which only differed from the target sentence in terms of the verb. However, the verbs in the distractor sentences were close in meaning to the target sentence verb, as shown below.

Target sentence: The man strolled through the park

Distractors: The man sauntered through the park
The man walked through the park

The results showed that the comprehension times for stimulus sentences are lower than the image-generation times. However, the probability of recognising a target sentence is much higher in the image-generation condition than the comprehension condition. This implies that Ss access information in the image generated for each sentence which allows them to distinguish between a target sentence and the semantically close distractors. Of course, the results are open to more than one interpretation because the difference in recognition probablities between the two conditions could be due to the difference between comprehension and image-generation times. To take account of this possibility, the experiment was repeated in a modified form.

Experiment 2 Subjects in condition 1, rather than just read the stimulus sentences were asked to categorize them according to whether the actions they described involved an object as an instrument of the action. For example, sentence S1 would be classed as NO INSTRUMENT, whereas S3 would be classed as INVOLVING AN INSTRUMENT. The categorization RTs were measured. The rest of the

S3 The chef chopped the vegetables.

experiment was the same as Expt. 1. The results showed that the categorization times were slightly longer than the image generation times, but the recognition probabilities were higher for the image generation condition than the categorization condition. Thus, the results seem to support the conclusions derived from the first experiment.

These experiments seem to imply that Ss access knowledge via images which allows them to differentiate semantically close verbs. However, it would seem that Ss do not access this knowledge in the normal process of comprehension as evidenced by the poor recognition performance in the comprehension condition. This can be construed as evidence against the involvment of imagery in language understanding, but if the I system takes longer than the P system to process an input then the I system may be used as a back-up processing capacity which is only resorted to when P system processing has failed, or proved inadequate. Evidence for this notion comes from work on the comprehension of metaphor which shows that people use imagery to work out the meanings of figurative inputs [Note 1]. The experiment below provides further support for this idea.

Experiment 3 - Subjects had to judge whether stimulus sentences were semantically acceptable or not and the decision RTs were measured. The sentences were constructed so that they would be easy to classify, in either direction, or difficult to classify leading to a bi-modal distribution in decision times. The same sentences were used in an image generation task and the response times were measured. As expected, the results showed a strong bi-model distribution for the decision RTs; Ss tended to be either quick or slow at making judgements. The important findings, however, were that (i) there was no significant difference between the image generation times for fast-RT sentences and slow-RT sentences, and (ii) the correlation between image generation time and decision RT was high for slow judgements, but low for fast judgements. These results imply that when Ss have difficulty in judging the semantic acceptability of a sentence they base the judgement on image coded data rather than knowledge stored in propositional form which is the case for more straightforward judgements. The evidence presented in this section gives weight to the theory that image processing is an important component of language understanding. Specifically, the data show that (i) image codes represent knowledge not directly available to the P-system, but which needs to be accessed in certain comprehension tasks, and (ii) image processing is often employed to solve linguistic problems.

5 Building a complete Language Understander

To build a flexible language understander with powerful problem solving capabilities it is necessary to augment existing P-system parsers by adding image generation and processing routines. The I-system and P-system would function in parallel, with the latter producing a skeletal parse adequate for some comprehension tasks. The I-system would generate a richer (semantically) parse, but over a longer time course. When the P-system output satisfies the requirements of the comprehension task image processing is terminated, uncompleted, and the understander passes on to the next input. If, on the other hand, the output from the P-system is inadequate for the comprehension task, then the richer I-system output is used to solve the problem. At any processing stage after generation, image codes can be inspected and potentially translatable elements of a code can be passed over to the P-system in order to generate a verbal response. The I-system has no direct output mechanism.
Kosslyn's computer simulation of his I-system theory [11] provides a good basis for a working image processing component which can be knitted

into a P-system parser, although it needs to be extended to take account of dynamic images and event simulations. It is not clear how his system would generate an image of a man strolling, as he does not specify how it is possible for components of an image to move relative to each other in a meaningful way. To implement an event simulation of sentence S1 it is necessary to access and run over time an ordered sequence of "key" image frames which represent the actional component of the verb 'stroll'. If you imagine a continuous motion sequence to be broken down into a series of static images, then the key images are those which have a high information content in that they distinguish between the meanings of different verbs. Such images might correspond to a discontinuity in the movement sequence, or a change of direction of motion of an image element. These sets of key image frames should be abstract enough to take different objects/agents of an action as the content of the image. The actional component of the verb within the P-system would be linked to its corresponding key image frame sequence within the I-system, and when the actional component is accessed during parsing the key image frame sequence would be run in the I-system to generate the event simualtion. At the same time, the other P-system elements accessed during parsing, corresponding to the arguments of the verb, would activate object images to form the content of the event simulation. These object images would be interpreted by the key image frame sequence to produce the dynamic image. These ideas have only been discussed in the context of representing the verbs of movement, but they could be generalised to provide a basis for generating the full range of dynamic images and event simulations.

6 References

[1] Paivio, A. Imagery and Verbal Processes. Holt, Rinehart and Winston, New York, 1971.

[2] Schank, R.C. Identification of conceptualizations underlying natural language. In R. Schank and K. Colby (eds.), Computer models of thought and language. Freeman, San Francisco, 1973.

[3] Wilks, Y.A. Grammar, Meaning, and the Machine Analysis of Language. Routledge, London, 1972.

[4] Rumelhart, D.E. and Levin, J.A. A language comprehension system. In D.A. Norman and D.E. Rumelhart (eds.), Explorations in cognition. Freeman, San Francisco, 1975.

[5] Norman, D.A. and Rumelhart D.E. Explorations in cognition. Freeman, San Francisco, 1975.

[6] Kosslyn, S.M. Image and Mind. Harvard University Press, Cambridge, Mass., 1980.

[7] Boggess, L.C. Computational interpretation of English spatial prepositions. Unpublished Ph.D. dissertation, Computer science department, University of Illinois, Urbana, 1978.

[8] Waltz, D.L. Generating and understanding scenen descriptions. In Joshi, Sag and Webber (eds.), Elements of discourse understanding. Cambridge University Press, London, 1980.

[9] Paivio, A, and Begg, I. Imagery and comprehension latencies as a function of sentence concreteness and structure. Research Bulletin No. 154, department of Psychology, University of Western Ontario, 1970.

[10] Thorndyke, P.W. Conceptual complexity and Imagery in Comprehension and Memory. Journal of Verbal Learning and Verbal Behavior, 14, 359-369, 1975.

[11] Kosslyn, S.M., and Shwartz, S.P. A simulation of visual imagery. Cognitive Science, 1, 265-295, 1977.

Reference notes

Note 1. Slack, J.M. Metaphor Comprehension A special mode of language processing? Paper presented at the 18th. annual meeting of the Association for Computational Linguistics, Philadelphia, June 1980.

METAPHOR AS NONLITERAL SIMILARITY

Odella E. Schattin and Hartvig Dahl

INTRODUCTION. In many current theories, metaphors are thought of as covert or implied nonliteral comparisons, which are comprehended by reference to their underlying literal comparison statements, i.e. similes. This view can be traced back to Aristotle, and is evident in more recent accounts of metaphor comprehension (e.g. Kintsch, 1974; Miller, 1979; Searle, 1979).

This prevalent theory of metaphors is challenged by that of Ortony (1979a, 1979b, in press). While Ortony agrees with the notion of metaphors as implied comparisons, he contends that this "reduction" theory does little to explain the comprehension process involved in understanding metaphors. It is misleading, Ortony argues, to understand a metaphor in terms of its corresponding simile because the comparison referred to in the simile itself is often metaphorical in nature. In his words:

> "... the difference between the metaphor and its corresponding similarity statement is not that one is metaphorical and the other literal; the difference is that one is an indirect statement whereas the other is a direct one."
>
> (Ortony, 1979a, p. 177)

He proposes an alternative theory based on Tversky's (1977) contrast model which utilizes feature matching.

According to Ortony, the comprehension of a metaphor involves assessing the attributes or features which its topic and vehicle have in common. What is required for a good metaphor is that these shared attributes be more highly salient features of the vehicle term than of the topic term. This salience imbalance is one important mechanism which is involved in metaphor comprehension.

Another important aspect of Ortony's account is that the attributes shared by the topic and vehicle need not be identical. They need only be similar to each other. This similarity between the shared attributes can be, and often is, metaphorical in nature. Thus, another major concept in Ortony's theory is that of metaphorical or nonliteral similarity.

The present experiment is a preliminary study intended to test this notion that the basis for a good metaphor is nonliteral similarity. It is expected that the judged goodness of a metaphor will depend on two factors: a perceived similarity between the statement's two components, and the characterization of this similarity as nonliteral. Statements whose components are perceived to be literally similar will be judged low in quality, as will statements whose components are exceedingly dissimilar. It is also hypothesized that similarity and literalness are different, although related, dimensions.

METHOD. Design. The main part of the experiment involved obtaining ratings on three dimensions of 25 comparisons in the form of "Topic is like Vehicle", chosen in pretesting to represent the full scale range of "goodness" on a 9-point scale. The 25 statements ranged from 4 to 10 words in length with a mean length of 6 words.

Procedure. Six subjects, three males and three females, were asked to rate the statements first on similarity, then on literalness and finally on goodness or aptness. The order of tasks remained constant across subjects, but the statements were randomized differently for each of the three scales, and these randomizations were different for each subject.

RESULTS. The results indicate that the raters used the three scales reliably; the mean Coefficient Alpha was .82. Furthermore, the 25 statements used in the study provided a representative sample which extends over the range of each scale. Thus the implications pointed out below (see Discussion) are backed by a set of reliable scales and by metaphors which distribute uniformly along these scales.

Figure 1 is a scatterplot of the relationship between literalness and goodness. There is a significant negative linear correlation between these two scales ($r = -.67$, $p < .0001$).

As can be seen in Figure 2, although there is no significant linear correlation between similarity and goodness, there is a significant ($r = .60$, $p < .006$) curvilinear relationship. In fact, a good fit was obtained using a quadratic relationship between similarity and goodness, as represented by the equation underneath Figure 2.

While predictions of goodness from literalness were quite successful, they do not take into account the interaction of literalness with similarity. This interaction is evident in Figure 3, which is a scatterplot of the relationship between similarity and literalness. This figure indicates that at low values of literalness, similarity is only marginally related to literalness but at high values of literalness, similarity is rather well predicted by literalness.

As it turns out, this variance along the similarity dimension can be used to further improve our predictions of goodness (G). The best fitting quadratic relationship incorporating both similarity (S) and literalness (L) is represented by the following equation:

$$G = 1.86 + .74L - .16L^2 + .77S$$

Using this model, the predicted goodness ratings were plotted as a function of the observed goodness ratings. The resulting scatterplot can be seen in Figure 4. There is a highly significant positive correlation ($r = .92$, $p < .00003$) between the values generated by the model (predicted goodness) and those generated by the raters (observed goodness), thus accounting for 84% of the variance in the goodness ratings.

DISCUSSION. Our results support Ortony's rejection of the "reductionist" viewpoint of metaphor comprehension. We have shown that even statements traditionally referred to as "similes" vary in the degree to which their components are perceived as literally similar. Thus, the "reductionists'" claim that a metaphor must be converted to its underlying "simile" in order to be understood, is not well founded. Instead, we offer an alternative theory incorporating the notion of nonliteral similarity as the basis for good metaphoricity.

A viable interpretation of our model requires examining each component of the best fit quadratic equation:

$$G = 1.86 + .74L \quad .16L^2 + .77S$$

Note that at low levels, the literalness components become positive and the relationship becomes additive. As we have already noted in our discussion of Figure 3, at low levels of literalness, similarity can vary over a large range. Thus, at these low levels of literalness, if similarity is low then the goodness of a metaphor will also be low. If, on the other hand, while literalness is low similarity is high, the resulting goodness value will also be high. So, we have arrived at a quantitative illustration of nonliteral similarity as the best predictor of a good metaphor.

REFERENCES

Kintsch, W. The representation of meaning in memory. Hillside, N.J.:Erlbaum, 1974.

Miller, G. A. Images and models, similes and metaphors. In A. Ortony (Ed.) Metaphor and thought. Cambridge, England:Cambridge University Press, 1979.

Ortony, A. Beyond literal similarity. Psychological Review, 1979a, 86 (3), 161-180.

Ortony, A. The role of similarity in similes and metaphors. In A. Ortony (Ed.) Metaphor and thought. Cambridge, England:Cambridge University Press, 1979b.

Ortony, A. Understanding metaphors. In D. O'Hare (Ed.) Psychology and the arts. Hassocks, England:Harvest Press, in press.

Searle, J. Metaphor. In A. Ortony (Ed.) Metaphor and thought. Cambridge, England:Cambridge University Press, 1979.

Tversky, A. Features of similarity. Psychological Review, 1977, 84, 327-352.

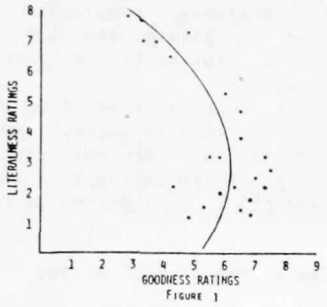


FIGURE 1

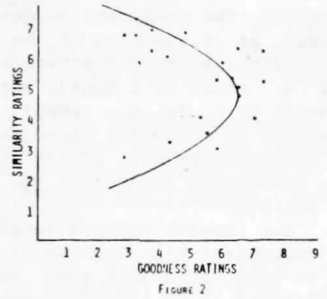


FIGURE 2

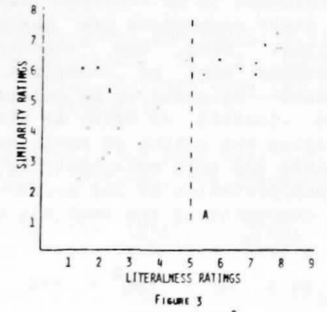
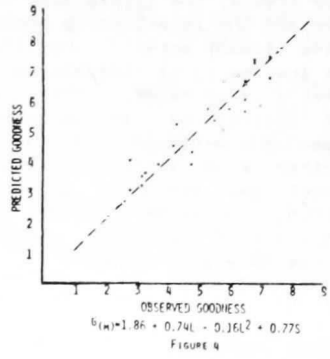


FIGURE 3

$G = -4.80 + 4.77S - .51S^2$



$G_{(M)} = 1.86 + 0.74L - 0.16L^2 + 0.77S$

FIGURE 4

A THEORY OF INTELLIGENCE

by: P. J. vanHeerden
POLAROID RESEARCH LABORATORIES
750 MAIN STREET, CAMBRIDGE, MASSACHUSETTS
02139

The basis of this theory of intelligence, in man or animal, is a model. The model is a computer, "the brain", with input channels and output channels. Shannon, in his theory of information[1] has shown that all information channels can be presented in the form of binary time series, sequences of ones and zeros only, as for instance 101101..... This is so because any record of events or things in the real world-be it photographs, television tapes, audio tapes, written or printed material, or any other imaginable record contains only a finite amount of information. Therefore, the model of intelligence can be a finite state digital computer, having only binary time series as inputs and outputs.

A further choice will be made in the model. It will be assumed that the program, that is the set of instructions given to the computer to make it into a model of intelligence, is independent of the size of the computer. This means that the model for intelligence in man, mouse, octopus or ant is the same in principle, although the amount of information processing needed to model human intelligence is clearly much larger than in the model of an ant.

It is almost silently implied that the program for the computer, to make it into a model of intelligence, is completely independent of the "actual meaning" of the information being processed. In a binary computer, the machine can only see the difference between the symbol 1 and the symbol 0, and nothing more.

If a man is hungry he will try to satisfy his hunger by eating. That is intelligent. The general principle of intelligence will be considered the generalization of that statement: "To be intelligent is to try to satisfy one's personal drives". However, one could not understand man, or social animals, without assuming the existence of several independent social drives-as for instance the parental instincts-which may very well overrule the drive of hunger in specific situations. If one is hungry, it may still be more satisfying to feed one's child than to eat one self. Man's spectrum of social drives has been very well represented for instance in the work of the social psychologist William MacDougall[2])

The drives will be represented in the model as a number of input time series to the computer. To satisfy one's drives one has to act, operate on the outside world, in some way. One may for instance ask where to find a restaurant, one may walk to it, one may pick up a menu. It will be assumed that intelligence results in commands to the muscles, as to the tongue in speaking, the legs in walking, the hand in holding the menu. These commands will be represented in the model as the output binary time series of the computer. There are, besides the drives, a second type of input time series needed for intelligent action. These are the channels which carry the information from the senses, as the eyes and ears, without which intelligent action would hardly be possible.

We now have all the components of the model of intelligence: a digital computer, with two types of binary input time series, and one type of binary output series. Let us arrange them in a diagram:

Model of Intelligence

	Past	Present	Future
Human Drives (as hunger, received from the body by the brain)	$H_1(1), H_1(2), H_1(3), \ldots$	$H_1(n)$,	$H_1(n+1), \ldots$
	$H_2(1), H_2(2), H_2(3), \ldots$	$H_2(n)$,	$H_2(n+1), \ldots$
	$H_3(1), H_3(2), H_3(3), \ldots$	$H_3(n)$,	$H_3(n+1), \ldots$
	----------	------	----------
	----------	------	----------
Human Senses (information flow from eyes, ears etc. to the brain)	$S_1(1), S_1(2), S_1(3), \ldots$	$S_1(n)$	$S_1(n+1), \ldots$
	$S_2(1)$, ----------	------	----------
	----------	------	----------
	----------	------	----------
	----------	------	----------
Commands to muscles (the information flow in the nerves from the brain to hand, tongue or foot)	$C_1(1), C_1(2), C_1(3), \ldots$	$C_1(n)$,	$C_1(n+1), \ldots$
	$C_2(1), \ldots$	$C_2(n)$,	$C_2(n+1), \ldots$
	----------	------	----------
	----------	------	----------
	----------	------	----------

H, S, C are all binary functions at specific intervals of time t: for instance $H_1(t=1)=1$, $H_1(t=2)=0$, $H_1(t=3)=0$, etc. The intervals $\Delta t=1$ may be 1 second, 0.1 second or the like.

[It is well known that the observing brain also acts on glands in the body to prepare it for action: the sight of food stimulates the digestive juices, the sight of fearsome things puts adrenalin into the blood, etc. Although this is not muscular action, the function, logically speaking is so similar to muscular action that it will be incorporated in the model as channels C_k, C_{k+1} etc.]

All we have to do now to finish the model of intelligence is to add a computer program that constructs, from the series $H_1, H_2 \ldots, S_1, S_2 \ldots$ and $C_1, C_2 \ldots$, the output series $C_1, C_2 \ldots$. The aim of the program is to make the future digits of H_1, H_2 equal to zero. H=0 will be assumed a satisfied drive, H=1 an active drive. It must be clear that one cannot expect a strictly causal relation between our future desires H and our present muscular activity. When one is out in the middle of a desert, it will be hard to come up with an immediate action to satisfy one's hunger or thirst. One can only ask for the best one can do, that is as many 0's as possible in the future of the H's, on the basis of one's limited capabilities.

Our program needs a theory of prediction. It is clear that such a theory should exist. People do learn to satisfy their needs by their actions. We call that skill, experience, insight, intelligence, "know how" or wisdom. One is not continuously succesful by just sheer luck. We also believe, on the basis of modern science, that the learning process takes place by information processing in the brain, analogous to what digital computers do. Does a theory of prediction exist? I will propose one[3]) to you. It is my view that such a theory, through the work on information theory, by Shannon and others, has become almost selfevident. The knowledge, that one can work with binary time series only, and arrive at a completely general theory, has certainly eliminated all mathematical difficulties. The only problem is scientific: to propose a theory that conforms to the intuitive expectations we have, as intelligent beings, about the future.

To begin, assume a binary time series f(t), and f(1)=1, f(2)=0, f(3)=1, f(4)=0, f(5)=1, f(6)=0. What is f(7), f(8), ers.? Let me write down the pattern: 101010.... One has the intuitive expectation that it will continue with 1010. Can we now find the reason for this intuitive expectation and formulate it as a general theory for all possible situations of binary time series?

Suppose one gives a specific time series to a group of scientists, experts in analysing information:
10110100101010001110.....
Let the time series be a long one, say a million digits, and one asks them to predict the future of the series. Naturally they will want to know what the time series represent, so they have some clues: is it the binary encoding of some spoken, or written language, the song of a bird, a television program? But they are told that a new born baby does not have the privilege of knowing what the world is all about. It receives signals from its eyes and ears and stomach and does not know what they mean. But in five or ten years it grows up into an intelligent being which knows a lot. Why should a group of accomplished scientists be entitled to more? So the experts start analysing the time series by all the means they have available and end up confessing that the series look very random to them, and therefore there is little to predict. One then gives them the clue that it may represent the first million digits of $\sqrt{\pi}$, written in binary digits. They try it and it fits perfectly. They will then have a strong intuitive expectation that also the future digits of the series will be represented by $\sqrt{\pi}$ Are they certain that this expectation will turn out correct? No, nothing is certain in the real world. It is simply the best prediction they have as rational beings on the basis of the available information.

This example shows the way to the general formulation of a theory of rational prediction. There would have been no reason to associate the time series with $\sqrt{\pi}$, because $\sqrt{\pi}$ would have been just one hypothesis out of a billion others one could come up with. But once a hypothesis has been tried and shown to conform to the presently given series, we have a strong intuitive confidence that this makes a valuable prediction. So the theory says: since all hypotheses are apriori equivalent, choose a handy package of hypotheses which is readily formulated, easily stored and efficiently tested. There exists one such handy package. It is: "compare the present time series with all its own pasts." For a million digit series, we have a million independent hypotheses without any effort. There is no other set of hypotheses I have been able to think of that matches this one in efficiency. I therefore hypothesize that this is the system used by all living creatures endowed with intelligence, great or small.

There are a few things to be specified before the theory is complete: The first one is that one must allow a percentage of error in the comparison of the present time series with the hypotheses. Otherwise one may find that not a single hypothesis is confirmed. The second one is that one can demand a match of the time series with the hypothesis, not over the whole series but only over an unspecified period back into the past. Then, if one has several hypotheses matching the time series, one over 10 digits one over 100 digits say, of the most recent past, one selects as the best prediction that one hypothesis which has matched the time series over the longest most recent past.

There is one more supplement to the theory which seems desirable. In our model, there is not just one time series, but a large number of series. We have presented the past, of Drives, Senses and Commands to muscles information as a two dimensional tableau of ones and zeros. To predict, we compare the present tableau not just with the tableaus we obtain by going 1,2,3 or more digits to the left. Instead, we also allow movements up and down. We investigate the match of the present tableau with all

other tableaus we obtain by going any number of steps to the left combined with any number of steps up or down.

The theory says that the brain is programmed to test n hypotheses, where the information stored by previous events in the life of the intelligent individual is n bits. It then selects the one hypothesis which conforms best to the present situation. For humans, this may require the testing of 10^9 to 10^{10} hypotheses. And don't forget, this task is performed not just once, but again and again, every time interval Δt of 0.1 seconds may be. Clearly that is an enormous computational task! There must therefore be short cuts.

To finish however, we must consider two situations. The first one is the case that the prediction leads to an increased satisfaction of the drives, an increased number of zeros. That is the easy one. The brain simply produces the corresponding commands to the muscles from its memory. This is as in the pleasant situation where one drives home after work in light traffic. One knows what to do. The second possibility is that the prediction is one of increased anxiety. "One enjoys one's dinner at home and notices through the window one's creditors converge on the front door" Or: "one enters one's home and finds a boa constrictor in the living room" What is now the model for the automatic instruction to the muscles? These questions are not answered as readily, and clearly require more discussion.

April 24, 1981

References

1 C. E. Shannon and W. Weaver "The Mathematical Theory of Communication," The University of Illinois Press, Urbana 1949

2 W. MacDougall "An Introduction of Social Psychology," Barnes and Noble, New York 1960 (originally 1908)

3 P. J. vanHeerden "The Foundation of Empirical Knowledge," Wistik, Wassenaar, Netherlands, 1968

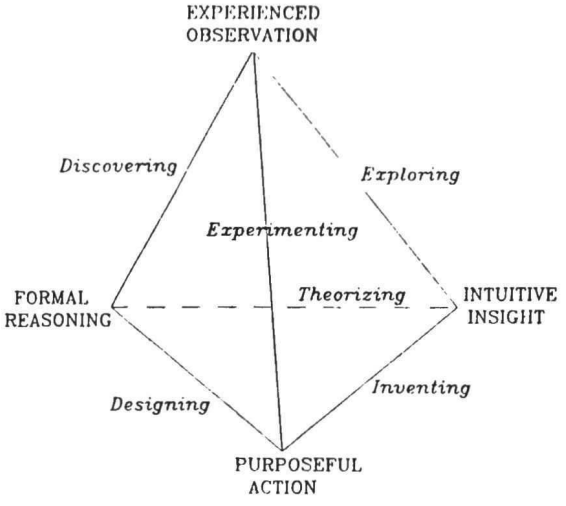
A TETRAGENIC FRAME FOR MODELLING UNMEDIATED KNOWLEDGE ACQUISITION*

Egon E. Loebner
Computer Reserach Center
Hewlett-Packard Laboratories

The question 'How does knowledge get into human heads?' has been asked many times before. In this paper the question is narrowed and sharpened two times. The first time, all knowledge mediated through direct contact with other human beings, as well as through various communication technologies, is exluded. Such knowledge has often been called immediate knowledge. In order not to prejudge the timing aspects of the acquisition process, it is called unmediated knowledge instead. This excludes all knowledge obtained through education as well as most learning activities. It thus restricts knowledge acquisition to processes through which people discover facts by themselves, invent their own solutions to problems and, without being told, shown or otherwise guided, figure out meanings of facts. Obviously unmediated knowledge cannot be experimented with apart from mediated knowledge without infraction of ethical codes. Unmediated knowledge acquisition is synonymous with knowledge generation. It is therefore referred to as gnomogenesis. The second sharpening of the above question comes from restricting it to gnomogenic events whose products are being reported for the first time. This restriction minimizes the probability that the knowledge had been obtained by mediation.

By singling out for study those gnomogenic events which are associated with recognized inventors, discoverers, explorers, designers, experimenters and theorizers, a more refined examination of gnomogenesis becomes possible. These six types of individuals are responsible for contributions in science, technology, art and politics. They often become singled out for special attention and, on occasion, their cognitive processes become scrutinized.

While one cannot sharply distinguish the six types of individuals, they can be associated with six cognitive processes



Model of Unmediated Knowledge Acquisition

responsible for their designations. These processes are inventing, discovering, exploring, designing, experimenting and theorizing. They are displayed, together with the four instrumentalities of gnomongenesis from which they can be derived, in the above figure.

Since the best documented cases of gnomogenic phenomena are found in the fields of science and technology, we begin the discussion with their products, the cognitive processes of discovery and invention. Discovery is a process of finding out something that was previously unknown. A classical example is a discovery of a new element in chemistry or a new particle in physics. After the discovery has been made and accepted by the scientific community the common belief includes not only belief in the existence of such an element or particle but also

*Based on material taught during the late 60's and early 70's at Stanford Universi
a graduate course entitled 'Introduction to the Heuristics of Invention and Disco

the belief that it existed before the discovery event, even though it had not been known prior to the event. It is important to note that its name is closely associated with the concept of uncovering. This implies the removal of an obstacle to observation. This obstacle is situational. The situation can be conditioned in the physical world or in the belief system the discoverer shares with his contemporaries. The discoverer removes the obstacle to perception and makes his act known. His peers confirm the new perceptions of the, now more easily accessed, object of discovery.

The act of invention differs from the act of discovery. It calls for contriving and fabricating something that had its origin in someones imagination. It becomes realized only after its idea became established in the mind of the inventor. When the act of invention is complete and the invention turns into fact, the common belief that the invention didn t exist before the act remains. While discovering deals with externalities, inventing deals with insight. It engages the more private aspects of the inventor's cognition. The idea for a wireless telegraph is a new composite with Hertz-discovered radiowaves replacing wire for the purpose of signal transmission over long distances. The idea preceded building of the device. When analyzing inventive cognition one finds introspection combined with the aim of outward demonstration of a product. Analysis of the cognitive process of discovery however indicates an outward looking aimed at an inner restructuring of situational understanding. A test, based on psycholinguistics, can verify that discovery and invention are not interchangeable. We speak of 'discovering the truth', but not of 'inventing truth'. We do say that 'a lie has been invented'. A 'discovery of a lie' could however be interpreted as a discovery of the existence of a lie, after it had been created. Thus we consider invention and discovery to be opponents in a similar sense that blue and yellow are opponents in the psychophysical space of color. This we symbolize by assigning them to two non-intersecting edges of a polyhedron.

It is necessary to exercise caution when dealing with so-called opponents. A full distinction cannot and should not be drawn between discovery and invention. While in some sense they are opposites in another sense they are not different. Here the concept of complementarity, introduced into quantum physics by Bohr for the purpose of escaping the dilemma of the particle-wave dualilty, can become a useful guide for the simultaneity of incompatibles. We plan to discuss this matter below.

A strong case for complementarity can be made for the second pair of gnomogenic processes: Experimenting and theorizing. That the former deals with matters external to a cognitive system, and the latter with those internal to it, need not be justified to a group of scientists. That the former swings between passive observation and active manipulation of externals and the latter between internal passive looking, literally intuiting, and internal manipulation of abstraact objects needs no demonstration to practitioners of the art of doing science.

The case for complementarity for the third pair of cognitions may be less obvious. Exploring deals with external observing of objects whose internal existence has been anticipated through internal visualization, while designing deals with internal manipulation that anticipates external manipulation of concrete objects.

Those familiar with lattice theory know that the six edges of a tetrahedron derive from forming logical unions between all of its four vertices. Equivalently the six gnomogenic functions could be derived by forming unions between four terms that suitably represent the labeled concepts at the vertices of the figure's tetrahedron. The product of two two-valued attributes will do that. The first attribute is MODE of cognition. Its value is perceiving, P, or atcualizing, A. The second is TENDANCE of cognition, signifying attendance of the environment or the self, with values E and S respectively. The tetragenic vertices of the tetrahedron become PE, AE, AS and PS in turn for top, bottom, left and right. The above simple but powerful axioms lead directly and systemically to a derivation of the four sources and six functions of unmediated knowledge acquisition. The high symmetry of this model's framework points to ease of manipulation and suggests some relationship between theories of cognitive science and those of particle physics.

What kind of technical meaning can we assign to the space bounded by the edges of the tetrahedron which is depicted in the above figure? Having labeled the edges with the six gnomogenic functions, paired into three opponents or complements, one can recognize such a space as belonging to psychophysics and psychometrics which, as Stevens points out, utilize the same scale to measure aspects of subjective responses (sensations, perceptions, judgments, etc.) and people (inventors, discoverers, etc.). Such space expresses relations between the subjective gnomogenic processes or between people assigned in accordance with the dominant function they exercise when they generate new knowledge. This space can also be utilized to chart the sequence followed by a single scientist, inventor, technologist, etc., mapping progression of his ideations after their articulation. Such graphs can display differences among such individuals and identify their strengths as well as weaknesses. They make possible recognition of dominant traits, such as observational power, reasoning power and the like.

Models similar to this tetragenic model have been used before. A pre-Socratic philosopher, Empedocles, formulated the four element model of the composition of ponderable matter. Only after 2000 years were the four elements of earth, water, air and fire replaced by the current theory of over a hundred chemical elements, called atoms. However, these are today subdivided into more primitive parts, and so on. Nevertheless the

original idea, that matter is to be thought of as composed of more elemental parts, still prevails. The tetragenic model of gnomogenesis should be considered in a similar light. It accounts for the psychophysical fact that perceptions, judgements and attitudes can be decomposed into more primitive elements. Another model is Carl Jung's attitudinal model of individual types. It contains three pairs of opponents that are not equivalent. The extroverted-introverted pair is major, and the thinking-feeling and sensing-intuiting pairs are minor. Jung's theory states that one major and one minor faculty always dominate the personality of an individual. This leads to an eight-fold typology of individuals. By "feeling" Carl Jung meant the "faculty of weighing and evaluating experience", which he thought as rational as the intellectual "faculty of thinking". By "sensing" he meant the "faculty of objective presentation" and by "intuiting" "an involuntary act that lacks judgment". Our tetragenic model differs from Jung's and other models interpreting intuition. This difference leads not only to a differing interpretation of its relationship to the other gnomogenic instrumentalities of cognition, but also points toward a need to investigate different experimental phenomena within the psychology of intuition.

Let us examine a few authoritative descriptions and definitions of intuition. Quinton distinguishes between two kinds of definitions of intuition: one ordinary and the other technical. According to the ordinary one, intuitions are expressed by making rapid and accurate assertions about matters of fact in circumstances where reliance on standard procedure is ruled out. This ordinary, nontechnical, sense has been adopted by experimental psychlogists such as Hebb and Westcott when designing their experiments to test for intuitive behavior in human subjects. Among the technical ones, the most familiar definition describes intuition as the power of obtaining knowledge which cannot be acquired by either inference or observation, reason or experience. There is little question that Peirce, one of the clearest thinkers and a great philosopher of knowledge, believed that the genetic endowment of man includes, besides animal instincts and everyday "common sense", most importantly, "in the cognitive domain a sense of the plausible regarding of the workings of nature". A remarkable articulation of the nature of intuition comes from Eaton, a virtually unknown American philosopher of knowledge, who wrote during the 20's. He stated: "What intuition gives us is a residue of knowledge, left over when all that is clearly conceived or sensed in the object (the individuality of a pebble picked up on the beach) has been analyzed away". Eaton was critical of Bergson's anti-intellectualistic philosophy which equated intuition with "pure awarness" and by setting it on its own feet and thus severing "rational thought from the non-conceptual medium into which and from which it flows". Eaton saw Bergson doing violence to the cognitive act in looking "at cognition abstractly in the very effort to fasten on that which is most concrete in it". This description is reminescent of that given by Einstein, who described Bergson's theory of intuition by saying that Bergson was sawing off the tree branch on which he sat. Eaton saw cognition as "a fusion of reason, sensation and intuition".

The psychophysical tetragenic model, which we propose for cognition, is an extension of Eaton's theory of knowledge, which recognizes as the three inseparable instrumentalities of knowledge: reason, sensation and intuition. We have added a fourth instrumentality, purposeful action, and also have renamed slightly the other three. With hindsight, that uses knowledge produced recently by molecular biologists and computer scientists, we suggest that intuitive insight perceives facts about plausible workings of nature indirectly through examination of the intuiter's own physiological structure which has been constructed from genetic plans supplied by preceeding generations. This explanation provides not only a plausible hypothesis as to the origin of a priori knowledge, but also to the mechanism by which this knowledge is being constantly enhanced. It also suggests the possibility of providing intelligent machines with the power of intuition if a way can be found for the machines to gain information through inner examination of their own structure, whether hardware or software.

Experienced observation is much more than pure sensation. The whole cognitive system is attuned to integrate perceptual regularities, which map the structure of the cognizer's environment into his own, as they are transmitted through his senses and higher level cognitive processors. As pointed out by N. R. Hanson, scientific observation is inseparably loaded with theory. It should be kept in mind that the experimenter swings between observation and contrived circumstance just as the theorizer wanders between insight into his subject matter and skillful use of his reasoning methodology. By experimenter's contrived circumstance is meant an elaborate experimental set-up, loaded with instruments so that the capture of the observed phenomenon is obtained under well understood, if not completely controlled conditions. The quantum physicist has taught us that pure observation without interference from the purposeful act of measurement violates the principle of uncertainty. This means that, in general, psychophysical phenomena should be mapped only in the interior of the tetrahedron. That is a consequence of Eaton's concept of fusion of all the instrumentalities during the cognitive act. A similar explanation holds for the nature of theorizing. Swinging between intellectual reasoning and intuition is a tenet of Bergson's theory of knowledge as well

Our model describes a special theory

of cognition which may eventually provide additional leads toward a more general understanding of cognitive functions. Its interdisciplinary approach draws on special knowledge in physics, psychophysics, psychology, biology, logic, computer science and philosophy. The provisional results point toward the possibility of generalizing physic's indeterminacy principle to six principles of indeterminacy for each of the cognitive functions formed from the four gnomogenic sources.

For psychology the tetragenic model points toward a better formalization of metapsychological approaches to cognition, where psychologists study the workings of psychologists as well as other scientists. We believe that phrases containing terms like "obvious", "self-evident" and "must" are direct pointers to intuitive insights of theorizers and inventors, when they use them during reports and descriptions of their accomplishments. We think that insufficient attention is currently given to intuition by cognitive science. Using up-to-date tools, studies could follow directions indicated already by Hovland's "invention of entirely new concepts" and by Wertheimer's "productive thinking".

For the field of machine intelligence intuition points toward the design of machines which have power to gleen knowledge about the intentions of their designers by intuiton, that is examination of their own hardware and software structures and fusing that knowledge with the knowledge obtained from the other three instrumentalities, which are already part of the state-of-the-art of computers.

SURREALISTIC IMAGERY & THE CALCULATION OF BEHAVIOR

Sheldon Klein, David A. Ross, Mark S. Manasse
Johanna Danos, Mark S. Bickford
Walter A. Burt & Kendall L. Jensen

Computer Sciences Dept.
Mathematics Dept.
Linguistics Dept., Music Dept.
University of Wisconsin
1210 W. Dayton St.
Madison, Wisconsin 53706

We present a model for human cognitive processing which assumes that a major component of the rules for calculating behavior are resident outside the individual, in the inherited, collective phenomena that anthropologists call 'culture.' Our model contains rules of behavior encoded in propositional structures such a frames or scripts, plus a method for calculating behavior by analogy. The propositional rules are transformed into situational state descriptions that are related by transformation operators which are also valid for transforming situations analogically. This kind of operator has been named, 'Appositional Transformation Operator,' or ATO (Klein 1977). The term refers to a theory which posits that the division of labor between propositional and appositional modes of reasoning is culturally determined (TenHouten & Kaplan 1973; Paredes & Hepburn 1976, 1977). The ATO's described by Klein (1977) are derived from the 2-valued, strong equivalence operator of mathematical logic, and from a 3-valued variant. ATO's, mechanically derived from propositional frame descriptions, can be used to calculate behavior by analogy, and in a way that evades many combinatorial problems associated with computation using propositional forms. Tasks of goal related planning can be handled in particularly efficient ways.

The model represents a radical interpretation and extension of ideas of Claude Lèvi-Strauss (1962, 1964-71). We suggest that many of the ATO's in a given society are encoded in the material and symbolic artifacts of its culture. It is not necessary to assume that human processors compute behavior in significantly better ways than contemporary computer models. Rather, it is the collective, inherited cultural environment that supplies the information and constraints that make calculation of behavior a relatively simple task. We suggest that surrealistic imagery present in myth structures encodes ATO's that control social behavior. Surrealistic dream imagery encodes both collective and individual ATO's. The archetypes of Jung, supposedly resident in a collective unconscious, can be interpreted in this model as ATO's in the form of surrealistic, iconic imagery. This imagery emerges independently and in parallel among members of similar social groups as a consequence of their social interactions. ATO imagery is both culturally inherited and spontaneously emergent: it aids in learning the rules of social structure, and emerges as epiphenomena in the minds of individuals as a consequence of the structure and repetition in their everyday interactions. The imagery may shift with social change and the replacement of generations. This reinterpretation of Jung limits similar archetypes to similar sociocultural groups, and makes unnecessary any genetic or metaphysical basis for the concept of 'collective unconscious.'

To explore the ATO thesis, we have constructed a computer simulation model of an artificial society, culture and universe- a world derived from Edwin Abbott's 19th-century fantasy, FLATLAND: A Romance of Many Dimensions (1884). The model is embedded in a new version of the Meta-symbolic Simulation System (Klein et al. 1976, 1977, 1979; Appelbaum 1976) now interpreted in PASCAL for the Terak and Apple microcomputers, and named, ' ||||| ,' (pronounced 'BAR BAR,' for short). We have built a frame-driven model that generates an opera, complete with textual, visual and musical output, all derived from the same semantic source. The rules are somewhat similar in form to those of Propp (1928) and are related to the type in Klein et al. (1976, 1977), plus an improvement: we have now implemented Propp's concept of multi-move tales (i.e. the satisfaction of goals through the calculation of events in parallel). Our output is an opera entitled, REVOLT IN FLATLAND: An Opera in Two Dimensions, produced in the form of a videotape (Klein et al. 1981). At each time increment during the generation of the opera, the complete state of the universe is recorded in a two or three dimensional, binary feature array. This encoding is a consequence of the implementation of the data structures in ||||| (BAR BAR). We may calculate an ATO to relate any two such states. If we save the entire series of ATO's that relate the sequential states of the opera, we have the basis for generating a surrealistic version. To accomplish this we interpret the ATO's, not as operators, but as state descriptions of the universe (they are in exactly the same format as the state descriptions they relate). They may then be used, in original sequence, to generate text, visual imagery and music. It is our thesis that this multi-medium encoding is a significant component of the material and symbolic artifacts of a culture. We suggest that what anthropologists call 'culture' serves as a repository for the analogical operators (ATO's) that make computation of behavior (rules of social structure) a feasible task for human automata. It is also possible to compute ATO's that relate ATO's. These, also, may be interpreted as state descriptions and used to generate a kind of surrealistic imagery in the form of text, visual images and music. At the highest levels, such ATO's can relate patterns behind the patterns that relate patterns of behavior; these ultimately relate all the domains of behavior possible in a society. Such ATO's may be encoded as myth systems, deities or as other forms of iconic imagery. Such a model provides a new level of mechanism for ideas of Lèvi-Strauss (Klein et al. 1980).

We are currently working on a videotape performance of a surrealistic version of our opera, REVOLT IN FLATLAND, to accompany the original. The calculation of plots by propositional reasoning is quite time consuming. The ATO's derived from a simulation generation can also be used to generate a new opera by analogy. All that is required is to specify a new set of initial conditions and the quantification of classes referenced by the ATO's. The derivation of a new opera that is an analogue of the old can take place with calculations no more complex than operations between feature arrays, carried out in chained sequence.

The formal properties of ATO's permit calculation of successor and atecedant states. Accordingly, it is possible to plan by analogy. This kind of planning is not limited to goal states that are realistically attainable in the modelled universe. Because ATO's are analogic, one can calculate the hypothetical requirements to realize an unobtainable goal, via analogy with goal attainment sequences that are possible. The textual, visual and musical realization of our operas may also take a surrealistic or fantasy form if the starting conditions and hypothetical goal states that are specified are made very deviant from what is implied by the rules of primary simulation model. We are also working on the generation of analogical variants of the opera in which initial, intermediate and terminal states are specified that could not be

Z

A loves no one, has no $, is married to B. B loves A, has no $, is unmarried to A. C loves A, has $, is unmarried.

W

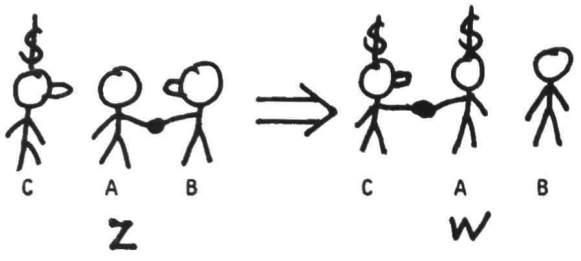
A loves no one, has $, is married to C. B loves no one, has no $, and is unmarried. C loves A, has $ and is married to A.

	La	Lb	Lc	$	Ma	Mb	Mc
A	.	0	0	0	.	1	0
B	1	.	0	0	1	.	0
C	1	0	.	1	0	0	.

	La	Lb	Lc	$	Ma	Mb	Mc
A	.	0	0	1	.	0	1
B	0	.	0	0	0	.	0
C	1	0	.	1	1	0	.

*ZW

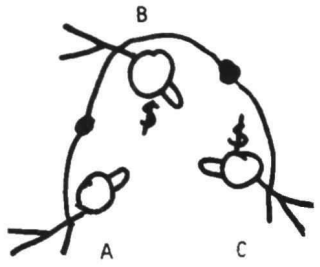
	.	1	1	0	.	0	0
	0	.	1	1	0	.	1
	1	1	.	1	0	1	.



*(*XY)(*ZW)

.	1	1	0	.	1	0
0	.	1	1	1	.	1
1	1	.	1	0	1	.

=



"surrealistic" interpretation

A loves B & C, has no $, is married to B. B loves C, has $, and is married to A & C. C loves A & B, has $, and is married to B.

If we then postulate a situation P,

	La	Lb	Lc	$	Ma	Mb	Mc
A	.	1	1	0	.	0	0
B	1	.	0	0	0	.	0
C	1	0	.	1	0	0	.

=

A loves B & C, has no $, and is unmarried. B loves A, has no $ and is unmarried. C loves A, has $ and is unmarried.

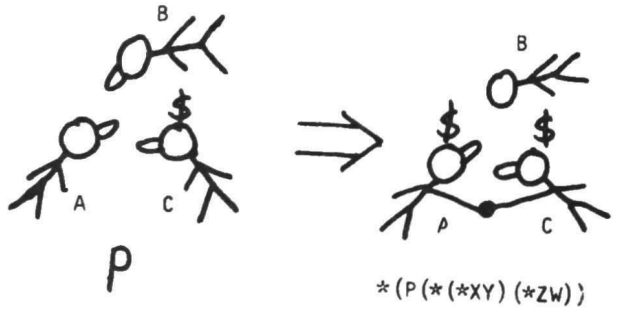
we can compute its successor state by analogy with the combined results of X⇒Y & Z⇒W by solving ((X :: Y) :: (Z :: W)) :: (P :: ?), where ? =

P((*XY)(*ZW))

	La	Lb	Lc	$	Ma	Mb	Mc
A	.	1	1	1	.	0	1
B	0	.	0	0	0	.	0
C	1	0	.	1	1	0	.

=

A loves B & C, has $, and is married to C. B loves no one, has no $, and is unmarried. C loves A, has $ and is married to A.



(P((*XY)(*ZW)))

Planning and Deviant Behavior

If a sequence of events, A, B, C, D, occur, then



If we wish to obtain a state E instead of D, without changing any of the ATO's, we derive, by analogy, a sequence leading to E by replacing A, B, C, respectively, with *A(*DE), *B(*DE), *C(*DE). If we wish to make a plan that specifies more than one goal state in the event sequence, we must alter some ATO's.

We suggest that the meaning of 'culturally defined behavior' is that members of a society plan in a way that minimizes the level and number of ATO's affected. It follows that deviant behavior may be interpreted as behavior that violates acceptable levels and numbers of ATO's. ATO patterns are part of the knowledge acquired by children. We posit that ATO patterns are encoded in multiple mediums of expression, both material and symbolic, and that they are the source of metaphor. It is this encoding that gives form to a culture, and it is this distributed presence in the environment that makes calculation of social behavior computationally feasible for the human mind.

References

Abbott, E.A. 1884. FLATLAND: A Romance of Many Dimensions. 2nd Ed., London. (1952. New York: Dover.)

Appelbaum, M.A. 1976. Meta-symbolic Simulation System (MESSY) User Manual. UW Comp Sci Tech Rept. 272, 169pp.

Klein, S. 1977. Whorf Transforms and a Computer Model for Propositional/Appositional Reasoning. Presented at the Applied Mathematics Colloquium, Univ. of Bielefeld, West Germany (Dec.); at the Computer Science Colloquium, Univ. of Paris-Orsay (Dec.); at a Joint Colloquium, Anthropology & Computer Sciences Depts., Univ. of California-Irvine (1978, March).

Klein, S., J.F. Aeschlimann, M.A. Appelbaum, D.F. Balsiger, E.J. Curtis, M. Foster, S.D. Kalish, S.J. Kamin, Y.D. Lee, L.A. Price. 1976. Simulation d'hypothèses émises par Propp et Lèvi-Strauss en utilisant un système de simulation meta-symbolique. Informatique et Sciences Humaines, No. 28, pp. 63-133.

_____. 1977. Modelling Propp & Lèvi-Strauss in a Meta-symbolic Simulation System. In Patterns in Oral Literature, H. Jason & D. Segal, eds., The Hague:Mouton.

Klein, S., D.F. Aeschlimann, D.F. Balsiger, S.L. Converse, C. Court, M. Foster, R. Lao, J.D. Oakley & J. Smith. 1979. AUTOMATIC NOVEL WRITING: A Status Report. In Text Processing/Textverarbeitung, W. Burghardt & K. Hölker, eds., pp. 338-412, Berlin: de Gruyter.

Klein, S., D.S. Kaufer, & C.M. Neuwirth. 1980. The Locus of Metaphor in Frame-driven Text Grammars. In The Sixth LACUS Forum 1979, W.C. McCormack & H.J. Izzo, eds, pp. 53-67, Columbia, S.C.: Hornbeam Press.

Klein, S., D.A. Ross, M.S. Manasse, J. Danos, M.S. Bickford, W.A. Burt & K.L. Jensen. 1981. REVOLT IN FLATLAND: An Opera in 2-Dimensions. Presented at the 5th Int. Conf. on Computers & the Humanities, May 17-19, Ann Arbor.

Lèvi-Strauss, C. 1962. La Pensèe sauvage. Paris: Plon. (English: 1966. The Savage Mind. New York: Basic Books.)

_____. 1964, 1966, 1968, 1971. Mythologiques, Vols. I-IV, Paris: Plon. (English: 1969, 1973, 1978, 1979. Introduction to a Science of Mythology, Vols. I-IV, New York: Basic Books.)

Paredes, J.A. & M.J. Hepburn. 1976. The Split Brain and the Culture-and-Cognition Paradox. Current Anthropology 17:121-127. (Discussion, 17:318-326, 503-511, 738-742; 1977. 18:344-350.)

Propp, V. 1928. Morfologija skazki. Leningrad. (English: 1968. Morphology of the Folktale. 2nd ed., Austin: Univ. of Texas Press.)

TenHouten, W.D. & C.D. Kaplan. 1973. Science and It's Mirror Image. New York: Harper & Row.

derived from the propositional rules, but which could be obtained by hypothetical analogy.

ATO's - Appositional Transformation Operators

ATO's relate situational descriptions in the form of feature arrays. A 2-valued version, essentially, is the strong equivalence operator of mathematical logic. If the interpretation of 1 and 0 are reversed, the operator is equivalent to non-carry, binary addition. The ATO is actually an array of bit operators defined as follows:

* a b	= c
0 0	1
0 1	0
1 0	0
1 1	1
. .	.

where '.' means 'does not apply,' making this specification 2-valued, with some augmentation.

The operator has the following properties:

*ab = "ATO"
*ab = *ba
*a (*ab)= b
*b (*ab)= a

e.g.

a		b
110	⟺	011
011		100
1.1		0.0

*ab

010
000 = "ATO"
0.1

A 3-valued variant is useful for state transitions where events emerge that were not present in the initial state. One can represent this with the 2-valued ATO, but the 3-valued variant is also useful. Again, a reversal of the interpretation of 1 & 0 yields an implementation as non-carry addition.

*	0	1	.	.0
0	1	0	.0	.
1	0	1	.	.0
.	.0	.	1	0
.0	.	.0	0	.

Examples of ATO's in Analogical Inference

Consider, first, some simple verbal analogies. A feature array referencing 'male,' 'female,' 'young,' 'adult,' 'love,' 'hate,' 'light,' 'dark,' is sufficient to formulate the following analogy:

X = Boy loves light : Z = Girl hates dark
Y = Woman hates light : ?

M	F	Y	A	L	H	Lt	Dk

where M = male, F = female, Y = young, A = adult, L = love, H = hate, Lt = light, Dk = dark

X = 10101010 : Z = 01100101 : *XY = 00000011
Y = 01010110 : ?

? = *Z(*XY)= 10011001 = Man loves dark

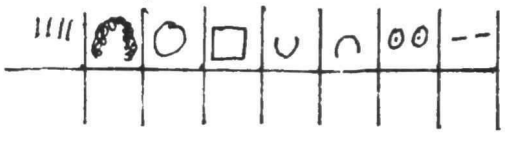
another example:

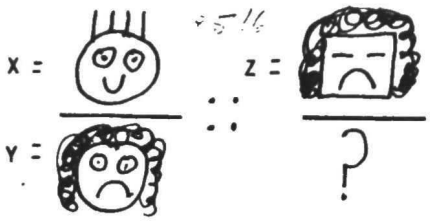
X = Man hates dark : Z = Boy hates light
Y = Woman loves light : ?

X = 10010101 : Z = 10100110 : *XY = 00110000
Y = 01011010 : ?

? = *Z(*XY)= 01101001 = Girl loves dark

The same method can be applied to visual analogies. For example, if the following set of visual features is used to create a pictorial analogy, the answer can be calculated using ATO's:





X = 10101010 : Z = 01010101 : *XY = 00110011
Y = 01100110 : ?

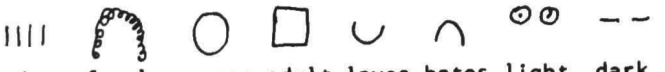
? = *Z(*XY) = 10011001 =



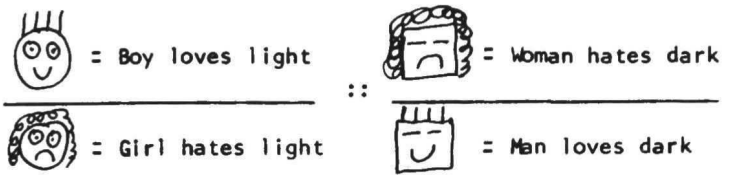
A visual interpretation of *XY might yield



If we give natural language interpretations to these visual features, such as,



male female young adult loves hates light dark



= Boy loves light : = Woman hates dark
= Girl hates light : = Man loves dark

ATO = *XY =  = sexless being, both old and young; indifferent to light and dark.

Complex analogies may also be computed, e.g.,

If ((X :: Y) :: (Z :: W)) :: (P :: ?) then ? = *P(*(XY)(*ZW)). Consider a complex example:

X	Y
A loves B, has no $, and is not married. B loves A, has no $, and is not married. C loves no one, has $ and is unmarried.	⇒ A loves B, has no $, is married to B. B loves A, has no $, is married to A. C loves no one, has $ and is unmarried.

Where La = 'loves A,' etc., $ = 'has money,' and Ma = 'married to A,' etc., the X Y states may be represented as follows:

X

	La	Lb	Lc	$	Ma	Mb	Mc
A	.	1	0	0	.	0	0
B	1	.	0	0	0	.	0
C	0	0	.	1	0	0	.

⇒

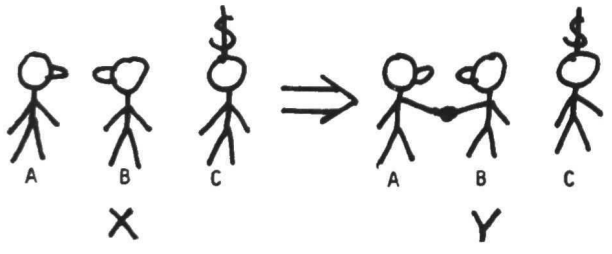
Y

	La	Lb	Lc	$	Ma	Mb	Mc
A	.	1	0	0	.	1	0
B	1	.	0	0	1	.	0
C	0	0	.	1	0	0	.

*XY

.	1	1	1	.	0	1
1	.	1	1	0	.	1
1	1	.	1	1	1	.

If we depict 'loves' as a nose pointing at the beloved (in between, if two loves), and if a noseless state means 'loves no one,' and if holding hands depicts 'married to,' and if a '$' indicates, 'has money,':



LEARNING WITH UNDERSTANDING

James G. Greeno
University of Pittsburgh

This paper is concerned with meaningful learning. Psychologists have distinguished between meaningful and rote learning (e.g., Katona, 1940; Wertheimer, 1945/1959) largely by providing examples that contrast the two phenomena. The work reported in this paper is an attempt to develop a more explicit and detailed theoretical analysis of the nature of learning that occurs with understanding.

I will consider learning situations in which new procedures and concepts are acquired for solving problems. Systems for learning procedures that have been analyzed previously are of two general kinds that I will call (1) direct learning and (2) analogical learning. I will describe a third kind of learning system in this paper that I call schematic learning.

In direct learning, examples are presented that show performance of the procedures that the learner is to acquire. Anderson et al (in press), Neves (1981), and Vere (1978) have studied processes of acquiring procedures that match the actions shown in examples or written theorems that correspond to inferential procedures. Processes in which fragments of procedures become integrated, forming new procedural concepts, also have been studied (Anzai & Simon, 1979; Larkin, in press; Neves & Anderson, in press), as have processes in which existing procedures are corrected, extended, or refined (Brown & VanLehn, in press; Goldstein, 1974; Neches, 1981; Sussman, 1975).

In analogical learning, a new procedure is acquired by mapping components of a known procedure to a new domain (Rumelhart & Norman, in press). The procedures that are transfered constitute new concepts that can be used to represent situations in the new domain.

In schematic learning, new procedures and concepts are formed in the framework of a general conceptual structure. A schema can provide a framework either for learning from examples or for analogical learning. I will discuss two examples that have been worked out in the form of running computational models that simulate salient aspects of student subjects' learning and performance. The first example involves learning to solve proof problems in geometry. This illustrates the role of a schema in learning from examples. The second example, which illustrates the role of a schema in analogical learning, involves learning procedures for multidigit subtraction in arithmetic.

Learning from an Example Proof

My first example is learning from the solution of a simple proof problem that is given early in a high-school geometry course. The problem and its solution are in Figure 1. I will discuss learning that can occur on the basis of this example problem, but first consider the problem in Figure 2, a problem that Wertheimer (1945/1959) discussed. Note that the solutions of these two problems are very similar in form. Three steps in Figure 2 correspond to the third step in Figure 1, but otherwise there is a simple mapping between the two solution proofs.

It might be expected that anyone who has learned to solve the problem in Figure 1 would also be able to solve Figure 2. It turns out that there is considerable variation in the success different students have with Figure 2 when it is presented as a transfer problem. A set of protocols on the problem in Figure 2 was obtained from students who had completed study of proof problems about line segments, such as Figure 1, and had begun to study properties of angles. Some students had no idea how to proceed. Others solved Figure 2 easily, and one even complained about having to solve "the same problem" so many times.

Consider the question: What enables a student to apply the knowledge acquired for solving Figure 1 to find a solution to Figure 2 easily? One hypothesis is that the procedures learned for solving Figure 1 were associated with general concepts that can be applied when Figure 2 is encountered. A version of this hypothesis has been implemented in a simulation program (see Anderson et al, in press, for a more detailed description).

The general structure that I postulate as the basis of transfer is a schema called Overlap/Whole/Parts. In this schema there are two components called "wholes," each of which is divided into parts, and a part of one whole is identical to a part of the other. I assume that in meaningful learning based on Figure 1, the Overlap/Whole/Parts schema is formed. Overlap/Whole/Parts has two subschemata, the Whole/Part structures that are included in the pattern. It is reasonable to assume that ninth-grade students have understood relationships of parts that form whole quantities for several years, and that they have some procedures associated with that schema. For example, they can add numbers associated with subsets to find the number in a superset, or subtract one part from the whole to form the other part.

The Overlap/Whole/Parts schema is formed as a combination of two Whole/Parts schemata, constrained so that a part of each "whole" component is shared with the other. Procedures that are attached to Whole/Parts are available in situations where the more complex structure is applied. In addition, some new procedures are also acquired and associated with the Overlap/Whole/Parts schema. For example, when the whole-components of the two substructures are equal, this enables the inference that the sums of their parts are equal, and when these sums are equal, the unshared parts are equal. (Learning of these procedures is based on Steps 4 and 5 in Figure 1.)

Two characteristics of the acquired knowledge are significant. First, the procedures that are acquired are defined on the components of the problem representations, which are the schematized versions of problems. This makes the procedural knowledge transferrable to other situations where the same schemata can be applied--for example, to problems such as Figure 2, if the system can learn to represent adjacent angles with the Whole/Parts schema. The second significant feature is that new conceptual entities are acquired when the schema of Figure 3 is learned. The organized structures of the wholes-with-parts are arguments of the new procedures, and thus function as cognitive units as a result of the learning that occurs.

Learning Subtraction Analogically

My second example involves the role of a schema in learning that is based on an analogy. This research has been done in collaboration with Lauren Resnick, who presents a companion paper in these proceedings. In our research, the learner does not construct the analogical mapping, as in the system that Rumelhart and Norman (in press) studied. Rather, the mapping between domains is presented in detail by an instructor. Performance of students indicates that this instruction leads to understanding of the procedure, and we consider the questions of what knowledge is acquired that constitutes this understanding and of how the

acquisition occurs.

The procedure that we have studied in this research is arithmetic subtraction. Children who were chosen to participate in the research had one of the subtraction "bugs" identified by Brown and Burton (1978). Examples of performance that involves bugs are in Figure 3. The first problem illustrates a "smaller-from-larger" bug, where the student subtracts the smaller digit from the larger one in each column, ignoring which is on top. The second and third problems illustrate a "don't-decrement-zero" bug, where borrowing from a zero does not include decrementing another number to its left or a change in the value of the zero digit.

The instruction that is given uses a procedure for subtracting with blocks. Different sizes of blocks represent different place values: small cubes for units, long (1 x 10) sticks for tens, flat (10 x 10) pieces for hundreds, and so on. Instruction occurs in three stages. First, a procedure is taught for subtracting with blocks. Second, there is a detailed mapping of that procedure to the procedure of subtracting with written numerals. Finally, the written procedure is made independent of the blocks.

The critical phase is in Step 2, where the correspondence between the procedures with the blocks and with the written numerals is made explicit. Each action in the blocks domain corresponds to an action in the written domain. For example, when a child removes a "tens"-block during a trade, the digit in that column is decremented by one, and when ten "ones"-blocks are added to the display, a small "one" is placed in the units column, indicating that ten has been added to that digit. This instruction can be considered as presentation of a component-by-component mapping between two procedures.

This instruction has been successful in changing children's performance, a form of debugging. Furthermore, children give us evidence that they have achieved significant understanding of arithmetic concepts and principles. One example was given by a student who had suffered from the smaller-from-larger bug. After instruction, this student was asked how the new procedure differed from the one the student used earlier. The student said, "I used to take the numbers apart; now I leave them together, ... and take them apart." We think that this shows that the student had achieved an understanding that the digits in one of the rows of the problem represent parts of a whole entity that is, that together they represent a number.

A second example was given by a student who had a don't-decrement-zero bug. After doing the problem: 403 - 275 correctly, including manipulations with blocks, the student was asked, "Do you know where the nine came from?" The student said, "It's nine tens, and the other ten is right here," and pointed to the small 1 that was written to the left of the 3 in the top number of the problem. We think that this shows that the student understood the principle of conservation involved in borrowing, that the numerals resulting from the borrowing procedure represent a quantity equal in value to reductions in another numeral.

Now consider the theoretical question: what knowledge is acquired in the instruction? Hypotheses about acquired knowledge should provide an explanation of the correct performance that results, as well as the evidence that students provide that they have achieved significant understanding. We will present two hypotheses. The simpler one uses an idea of schematic goals. The other hypothesis postulates that understanding of subtraction involves the Whole/Parts schema, the same structure to which we attribute understanding of the geometry problem considered above. The latter hypothesis has been implemented as a running program; the former is based on a suggestion by Robert Neches.

The hypothesis of schematic goals postulates that knowledge of the blocks procedure is organized in a way similar to Sacerdoti's (1977) system of hierarchical action knowledge, with higher-order actions providing a goal-based organization of lower-level actions in the procedure. Important goals for the blocks procedure include: (1) find an answer for each column; (2) if there are not enought blocks for a column, get some more; (3) if there are no blocks in a column where you need to get some more, get some blocks for that column. In the hypothesis of schematic goals, we assume that mapping instruction results in transferring the goals of the blocks procedure to the procedure with written numerals. We propose that these goals correspond to new cognitive units in the student's representation of subtraction with written numerals.

This organization can explain indications of understanding like those we presented earlier. The remark that the correct procedure "keeps the numbers together" is explained because the actions of Decrement-Top and Add-Ten are parts of the same general action. Similarly, the elementary actions Decrement, Makenine, and Addten are combined to form a larger structure, which could be the basis of the remark that "It's nine tens, and the other ten is right here."

The simulation that we have programmed is somewhat more complex than the hypothesis of schematic goals. Our reasons for implementing a more complex system were in protocols obtained as students learned about the procedures in he blocks domain. Instruction for this procedure involved a kind of discovery method, including questions such as, "Can you think of a way to get more blocks?" The student whose performance we tried to simulate showed several indications of understanding principles underlying the procedure without being shown the procedure. At one critical point, involving borrowing through zero, the student said "Ooh neat--Now I get it." We simulated the student's performance with a model in which adjacent digits are schematized as parts of a whole unit. Understanding of the part-and-whole relationship of adjacent digits enables the model to understanding borrowing through zero by co-ordinating a constraint of keeping a total quantity constant while adjusting the numbers of things in its parts. This is described in more detail in Resnick's companion paper.

References

Anderson, J. R., Greeno, J. G., Kline, P. J., & Neves, D. M. Acquisition of problem-solving skill. In J. R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, N.J.: Lawrence Erlbaum, in press.

Anzai, Y., & Simon, H. A. The theory of learning by doing. Psychological Review, 1979, 86, 124-140.

Brown, J. S., & Burton, R. R. Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science, 1978, 2, 155-192.

Brown, J. S., & VanLehn, K. Toward a generative theory of "bugs." Cognitive Science, in press.

Goldstein, I. P. Understanding simple picture programs. (Artificial Intelligence Laboratory Report TR-294). Cambridge, Mass.: Massachusetts Institute of Technology, 1974.

Katona, G. Organizing and memorizing. New York: Columbia University Press, 1940.

Larkin, J. H. Enriching formal knowledge: A model for learning to solve textbook physics problems. In J. R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, N.J.: Lawrence Erlbaum, in press.

Neches, R. Models of heuristic procedure modification. Unpublished doctoral dissertation, Carnegie-Mellon University, 1981.

Neves, D. M. Learning procedures from examples. Unpublished doctoral dissertation, Carnegie-Mellon University, 1981.

Neves, D. M., & Anderson, J. R. Becoming expert at a cognitive skill. In J. R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, N.J.: Lawrence Erlbaum, in press.

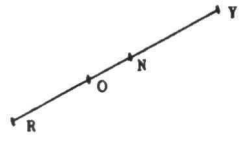
Rumelhart, D. E., & Norman, D. A. Analogical processes in learning. In J. R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, N.J.: Lawrence Erlbaum, in press.

Sacerdoti, E. D. A structure for plans and behavior. New York: Elsevier-North Holland Publishing Co., 1977.

Sussman, G. J. A computer model of skill acquisition. New York: Elsevier-North Holland Publishing Co., 1975.

Vere, S. A. Inductive learning of relational productions. In D. A. Waterman & F. Hayes-Roth (Eds.), Pattern-directed inference systems. New York: Academic Press, 1978.

Wertheimer, M. Productive thinking. New York: Harper & Row, 1945. (Enlarged edition, 1959)

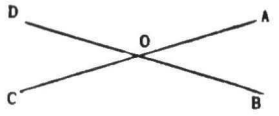


Given: RN = OY

Prove:: RO = NY

Statement	Reason
1. RN RO + ON	1. segment addition
2. OY = ON + NY	2. segment addition
3. RN OY	3. given
4. RO + ON = ON + NY	4. substitution
5. RO = NY	5. subtraction property

Figure 1



Given: $\overline{AOC}$, $\overline{BOD}$

Prove: $\angle AOB = \angle COD$

Statement	Reason
1. $\angle AOC = \angle AOD + \angle COD$	1. angle addition
2. $\angle BOD = \angle AOB + \angle AOD$	2. angle addition
3. $\angle AOC = 180^{o}$	3. def. of straight $\angle$
4. $\angle BOD$ 180°	4. def. of straight $\angle$
5. $\angle AOC$ $\angle BOD$	5. substitution
6. $\angle AOD + \angle COD = \angle AOB + \angle AOD$	6. substitution
7. $\angle COD = \angle AOB$	7. subtraction

Figure 2

$$\begin{array}{r} 3\ 2\ 7 \\ -\ 1\ 8\ 4 \\ \hline 2\ 6\ 3 \end{array} \qquad \begin{array}{r} 5\ 0\ {}^{1}2 \\ 3\ 0\ 6 \\ \hline 2\ 0\ 6 \end{array} \qquad \begin{array}{r} {}^{6}\not{7}\ {}^{1}0\ {}^{1}5 \\ -\ 2\ 3\ 9 \\ \hline 4\ 7\ 6 \end{array}$$

Figure 3

The Growth of Number Representation: Successive Levels of Schematic Learning

Lauren B. Resnick
University of Pittsburgh

In a companion paper*, Greeno has outlined a theory of learning as the successive construction of new schemata. In this paper we explore the effects of cumulating schemata on performance and competence in the domain of number knowledge. We begin with a brief consideration of the number representation assumed to be available to children before they have learned place value and decimal notation. Then we outline several stages in the acquisition of decimal place-value knowledge. Finally, we consider the implications of successive stages of number representation for a theory of the understanding of cardinality.

Early Number Representation

Varied databases on early counting abilities (Gelman & Gallistel, 1978; Fuson, in press), simple mental arithmetic (e.g., Groen & Resnick, 1977), and story problem solution (Vergnaud, in press; Carpenter & Moser, in press; Nesher, in press) have provided the basis for formal models of preschool and early school mathematical performances (e.g., Riley, Greeno, & Heller, 1978; Greeno, Gelman, & Riley, 1978; Briars & Larkin, 1981). These models converge on the following features of pre-decimal number representation:

1. Children possess an ordered string of "count words", linked by "next" and "backward next" relationships. Each position in the string has come to stand for a quantity. The string can be used to solve problems via counting. It can also be used as an analog representation of quantity. For example, the positions of two target quantities can be found and compared for relative "largeness" (Sekuler & Mierkiewicz, 1977).

2. Children can interpret small numbers in terms of a Part/Whole schema (Figure 1) such that any number can be interpreted either as a whole composed of two smaller numbers or as a part in a larger whole. The Part/Whole schema includes the constraint that the combined parts neither exceed nor fall short of the whole quantity. The schema has been shown to function in successful solution of certain story problems (e.g., those with the unknown in the first position) that younger children find very difficult. It seems likely that the schema also permits children to discover the complementarity of addition and subtraction, leading to a particularly efficient--and mathematically elegant--solution to subtraction problems. In this procedure, which has been observed in children as young as 7 years (Woods, Resnick, & Groen, 1975; Svenson & Hedenborg, 1979), children either count up from the smaller number or count down from the larger, whichever requires fewer counts.

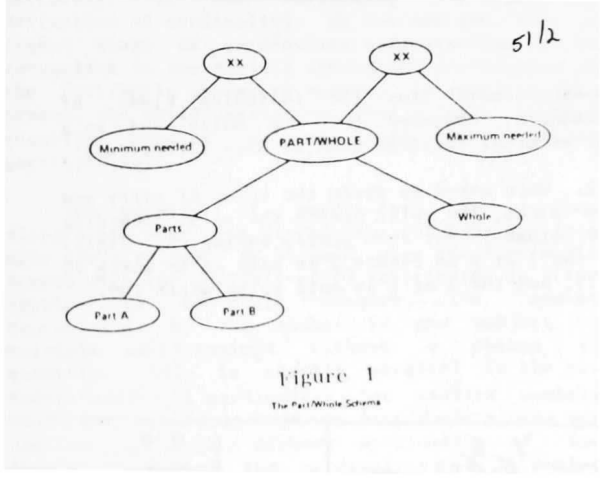


Figure 1
The Part/Whole Schema

Acquisition of Place-Value Schemata

In the course of learning the decimal number system, the string of count words is gradually reorganized to reflect an understanding that multidigit numbers are compositions of units and tens (later also hundreds, thousands, etc.). This is accomplished through successive elaborations of the Part/Whole schema. Several stages in the acquisition of this compositional interpretation of multidigit numbers can be identified in a program (MOLLY) that simulates the performances of a 9-year-old girl (Molly) as she acquired new knowledge through special remedial instruction in multidigit subtraction.

Four stages in MOLLY'S knowledge of place value can be distinguished:

Stage 1: Unique partitioning of multidigit numbers. At the earliest stage of place-value knowledge, MOLLY has a knowledge structure which organizes conventional information about the structure of multidigit written numbers (Figure 2).

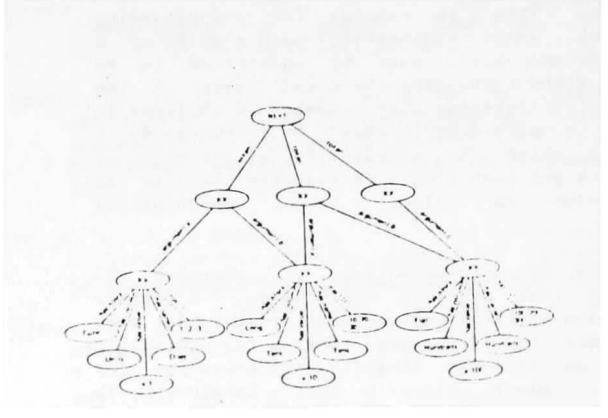


Figure 2

This structure identifies columns according to their positional relationship to each other (thus the centrality of the Next relationship). Attached to each column is a block shape (these refer to the shapes of blocks used in teaching base arithmetic), a counting string, a value, a column name, etc. Using the Next structure, MOLLY can:

1. construct a block display for any written number. This is done by identifying which column the number is in, finding the block shape that match that column, and displaying the number of blocks specified by the digit in the column.

2. *"read" a block display.* This is done by starting with the largest block shape, finding the counting string that matches it, enumerating the blocks using the appropriate string, and then iterating through successively smaller block shapes.

*Greeno, J. G. Meaningful learning. Paper presented at the meeting of the Cognitive Science Society, Berkeley, CA, August 1981.

3. interpret a written numeral as "x hundreds, y tens, and z ones."

4. compare block displays on the basis of the highest-valued block only (e.g., for 347 v. 734, compare only the hundreds blocks).

As long as the Next structure alone is used to interpret numbers, each written number can have only one block representation--a "canonical" representation, with no more than 9 blocks per column. This means that there is no basis for a semantic interpretation of the operations of carrying and borrowing. The next stage provides the earliest basis for this interpretation.

Stage 2: Multiple partitionings arrived at empirically. At this stage the Part/Whole schema is elaborated to include a special restriction, applied to two-digit numbers, that one of the parts be a multiple of 10. Application of Part/Whole permits multiple partitionings, and therefore multiple block representations, for any written number. Any specific partitions, however, must be arrived at through a counting solution. For example, to "show 47 with more ones," MOLLY first applies Part/Whole in a global fashion and then concludes that if the whole is to stay the same but more ones are to be shown, there must be fewer tens. It therefore reduces the tens by a single block. The schema is next instantiated with 47 in the Whole slot, and 30 in one of the Parts. The remaining Part is found by adding ones blocks and counting up until 47 is reached.

Two important concepts have been added to the number representation at this stage. First, the equivalence of several partitionings has been recognized. Second, the possibility of having more than 9 of a particular block size has been admitted. This is crucial for understanding "borrowing", where--temporarily--more than 9 of a given denomination must be understood to be present, without changing the total value of the quantity. Interviews with a number of children in addition to Molly make it clear that there is a stage in which the possibility of borrowing or trading to get more blocks is rejected because it will produce an "illegal" (i.e., noncanonical) display.

Stage 3: Preservation of quantity by exchanges that maintain equivalence. A further elaboration of Part/Whole appears at Stage 3, when MOLLY adds to its representation for multidigit numbers an explicit 10-for-1 relationship for adjacent block sizes. This knowledge is represented by a Trade schema (Figure 3) which specifies a class of legal exchanges among blocks.

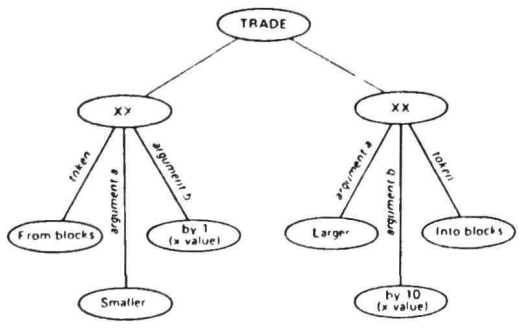


Figure 3

The schema specifies that there is a "from" pile of blocks, from which blocks are removed. This pile becomes smaller by one block. There is also an "into" pile of blocks that becomes larger by 10 blocks. The value of the blocks in the "from" and "into" piles is established by multiplying the number of blocks removed by the value of the block shape (as specified in the Next structure). Thus, when trades are made between adjacent block sizes, the schema specifies that both the "into" and the "from" values will be 10. Applied as an elaboration of the Part/Whole schema, the Trade schema allows MOLLY to conclude--without having to count up--that there has been no change in the whole quantity.

Stage 4: Application to written addition and subtraction. MOLLY also provides a theory of how the various levels of quantity representation discussed above can come to be applied to written numerals. MOLLY simulates the learning sequence achieved as a result of instruction that forced attention to the details of a mapping between the operations of written borrowing and those of block trading (Resnick, in press). Figure 4 shows the result of constructing this mental mapping: an abstraction that treats the two procedures as expressions of the same exchange principles, with analogous elements in the Trade and Borrow schemata. Evidence for this level of understanding

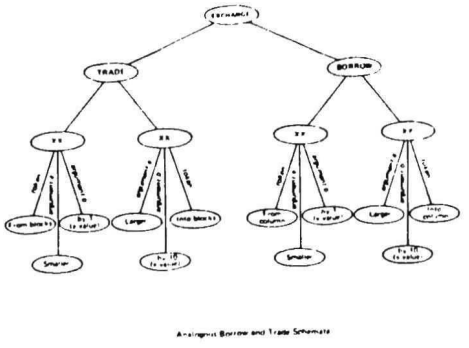
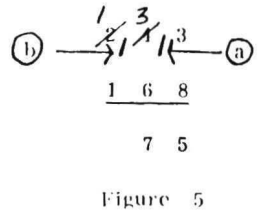


Figure 4

of number comes from the following kinds of performances, observed both in Molly and in a number of other children whom we have interviewed:

1. When asked to state the value of carry and borrow marks, the child states the value according to the column rather than simply naming the digit. Thus the 1 at a in Figure 5 is said to be worth 10 (not 1), and the 1 at b is said to be worth 100.



1 6 8

7 5

Figure 5

3 4 6 5

3 5 3 7

Figure 6

2. When asked to show the blocks that represent a given carry or borrow mark, the child selects blocks according to value. A ten-block is selected for a, a hundred-block for b, and so forth.

3. The child can construct justifications for the various markings in written subtraction. For example, Molly explained that in the subtraction problem shown in Figure 6 above, she had borrowed one thousand from the 7. When asked where she had put the thousand, she was puzzled at first, because there is no place where one can "see" 1000 in the markings given. However, she then said, "It is divided up. Nine hundred of it is here (indicating the hundreds column), and the other hundred is here (indicating the tens and the ones columns together). You see, this 9 is really 90 and this 1 is really 10 and that adds up to 100!" MOLLY is able to construct this explanation by first calling on the Exchange (Borrow) schema, which specifies that if a quantity of 1000 has been taken from a column it must be put into another column. Unable to find a column with 1000 put into it, MOLLY calls on the Part/Whole schema, sets the Whole slot equal to 1000 and looks for two Parts that add up to 1000. It finds 900, but cannot find a column with the 100 necessary for completing the Whole. MOLLY iterates through the Part/Whole schema again, this time setting the Whole equal to 100 and now finding 90 and 10 as the Parts.

An Interpretation of Cardinality

This characterization of children's developing number knowledge permits us to give a more precise psychological meaning to the understanding of "cardinality" than has heretofore been possible. Gelman and Gallistel (1978) included in their principles of counting a cardinality principle, which specifies that the final count word reached when a set of objects is being enumerated is the total number in the set--i.e., the set's cardinality. For the preschool child, who has not yet come to interpret quantity in terms of the Part/Whole schema, this is the only meaning of cardinality available. This criterion of understanding cardinality has been criticized, however, (e.g., Comiti, 1980) as too weak, and in particular as not reflecting the Piagetian definition of cardinality. We can now see that a higher stage of cardinality understanding can be recognized in the child's subsequent application of the Part/Whole schema to number. In applying this schema, the child understands that a total (whole) quantity remains the same even under variant partitionings.

The meaning of cardinality is further elaborated when the place-value schemata outlined here are acquired. At Stage 2, when the Part/Whole schema with the multiple-of-10 restriction is first applied to two-digit numbers, the amount represented by the number is now subject to multiple partitionings without a change in quantity. This is exactly parallel to the new understanding of cardinality for smaller numbers that was achieved when the Part/Whole schema was applied to them. Without application of the Part/Whole schema the cardinality of a number resides in the specific display set and the number attached to it through legal counting procedures. With Part/Whole, cardinality resides in the total quantity, no matter how it is displayed or partitioned.

The Trade stage of multidigit number representation represents yet a higher level of understanding of cardinality. Now it is recognized that cardinality is not altered by a specified set of legal exchanges. An analogy can be drawn with an earlier recognition of quantity as unchanged under various physical transformations (such as spreading out a display of objects-- the classic Piagetian test of conservation). However, the transformations produced under control of the Trade schema do in fact involve a change in the actual number of objects present. Thus, recognition that the value of the total quantity remains unchanged requires a level of abstraction concerning the nature of cardinality that was not required for earlier stages of understanding.

References

Briars, D. J., & Larkin, J. H. An integrated model of skill in solving elementary word problems. Unpublished manuscript, Carnegie-Mellon University, 1981.

Carpenter, T., & Moser, J. The development of addition and subtraction problem solving skills. In T. Carpenter, J. Moser, & T. Romberg (Eds.) Addition and subtraction: A developmental perspective, Hillsdale, NJ: Lawrence Erlbaum Associates, in press.

Comiti, C. Le concept de nombre chez l'enfant. Communication presented at the Quatrieme Congres International sur l' Enseignement des Mathematiques, Berkeley, CA, August 1980.

Fuson, K. An analysis of the counting-on solution procedure in addition. In T. Romberg, T. Carpenter, & J. Moser (Eds.) Addition and subtraction: A developmental perspective. Hillsdale, NJ: Lawrence Erlbaum Associates, in press.

Gelman, R., & Galistel, C. R. The child's understanding of number. Cambridge, MA: Harvard University Press, 1978.

Greeno, J. G. Gelman, R. & Riley, M. Young children's counting and understanding. Paper presented at the annual meeting of the Psychonomic Society, San Antonio, TX, November 1978.

Groen, G. J., & Resnick, L. B. Can preschool children invent addition algorithms? Journal of Educational Psychology, 1977, 69, 645-652.

Nesher, P. Levels of description in the analysis of addition and subtraction word problems. In T. Carpenter, J. Moser, & T. Romberg (Eds.) Addition and subtraction: A developmental perspective. Hillsdale, NJ: Lawrence Erlbaum Associates, in press.

Resnick, L. B. Syntax and semantics in learning to subtract. In T. Carpenter, J. Moser, & T. Romberg (Eds.), Addition and subtraction: A developmental perspective. Hillsdale, NJ: Lawrence Erlbaum Associates, in press.

Riley, M. S., Greeno, J. G., & Heller, J. I. Development of children's problem-solving ability in arithmetic. In H. P. Ginsburg, (Ed.), The development of mathematical thinking. New York: Academic Press, in press.

Sekuler, R., & Mierkiewicz, D. Children's judgments of numerical inequality. Child Development, 1977 48, 630-633.

Svenson, O., & Hedenborg, M. L. Strategies used by children when solving simple subtractions. Acta Psychology, 1979 43, 1-13.

Vergnaud, G. A classification of cognitive tasks and operations of thought involved in addition and subtraction problems. In T. Carpenter, J. Moser, & T. Romberg (Eds.), Addition and subtraction: A developmental perspective. Hillsdale, NJ: Lawrence Erlbaum Associates, in press.

Woods, S., Resnick, L. B., & Groen, G. J. An experimental test of five process models for subtraction. Journal of Educational Psychology, 67(1), 17-21.

The Role of Experiences and Examples in Learning Systems

Edwina L. Rissland+
Oliver G. Selfridge
Elliot M. Soloway*

Department of Computer and Information Science
University of Massachusetts
Amherst, MA 01003

Abstract

In this paper, we discuss the role of experiences and examples in learning systems. We discuss these issues in the context of three systems in particular: Rissland and Soloway's Constrained Example Generation (CEG) System, Selfridge's COUNT, and Soloway's BASEBALL.

1. Introduction

Examples and experiences, by which we mean concrete instances, situations or problems, are critical to any system, man or machine, that learns. Examples provide the basis from which generalizations, concepts and conjectures are made. They also provide the criticisms needed to refute and refine.

For instance, in Winston's learning program [Winston 1975], examples of the concept to be learned, e.g., an arch, and non-examples, e.g., "near misses", are the critical input from which his program builds a structural description of a concept. In Lenat's concept discovery program, AM, [Lenat 1977], examples help direct the discovery process, by providing evidence of the reasonableness and interestingness of new concepts.

+Supported in part by the National Science Foundation under grant IST-80-17343.
*Supported in part by the Army Research Institute for the Behavioral and Social Sciences under ARI grant MDA903-80-C-0508. Opinions expressed in this report are those of the authors, and do not necessarily reflect the views of the U.S. Government.

Examples are critical in human learning and discovery, whether it be in children [Hawkins 1980] or sophisticated adults. Lakatos [1976] gives a detailed exposition of the historically important Euler's formula; there, examples like various "monsters" (i.e., counter-examples) play a central role in concept refinement.

Thus, examples are grist for the learning process in a critical way. In this paper, we consider the role of examples in learning systems, in particular issues such as the richness of the base of examples upon which the system runs. Questions about the generation of examples are discussed in [Rissland 1980, Rissland and Soloway 1980, 1981].

2. Three Learning Systems

We now restrict our discussion to three learning systems to illustrate some general issues in learning. Briefly, the three systems work as follows:

1. BASEBALL possesses both high level schemas which describe the intentions of people in action-oriented competitive games and low level schemas which provide a common sense understanding of spatio-temporal events. From observations of activity in a baseball game, the system first interprets that activity, generalizes from these hypothesized rules, and finally accepts or rejects these rules based on their predictive utility. Rules for concepts like "out", "hit", "single" are learned.

2. CEG generates an example to meet posted desiderata by modifying known examples from its knowledge base of examples, its "Examples-space". In the version of the system being used to study learning, the system possesses several operators that can modify a given feature; its task is to explore both the space of examples and the space of modification operators not only to arrive at a solution -- a base example plus a sequence of modifications -- but also to gain experience in using the operators in order to allow later learning about the operators themselves.

3. COUNT possesses a repertoire of primitive number and string manipulation routines, such as "increment by 1", "move the pointer right by 1", and control routines "repeat", and "do N times", from which it is to build procedures, i.e., strings of primitives, and ultimately a "count" procedure to count the number of symbols in a string. The system learns by

solving problems posed by its user who acts as its teacher.

3. Examples and Learning

Each of the above three learning systems is provided experiences and examples upon which it bases its learning. The provider of these examples, in effect, acts as its teacher.

BASEBALL: the ensemble of observed games
CEG: the initial Examples-space, posed problems
COUNT: posed problems

Thus, the systems gain experience that ranges from a set of example games, to examples and problems, to just problems.

In each of the systems, there is a classic trade-off between the richness and size of the initial knowledge (not only of examples) and the amount and care of search that must be made for solutions and conclusions. The initial knowledge is used to control the size of the search space. The amount of search in the system varies:

BASEBALL: small
CEG: small-medium
COUNT: medium-large

Within CEG itself, for instance, the richer the initial Examples-space, the less care was needed to explore "adequately" the space of operator sequences. COUNT works with very little embedded knowledge and expends a large effort in search. BASEBALL generates a small number of interpretations and generalizations.

All of the systems make use of evaluation and judgement mechanisms:

BASEBALL: uses a teacher-specified threshold to accept/reject hypotheses based on their predictive utility
CEG: possessed by an explicit JUDGE module
COUNT: performed by the system for the ultimate counting task, and by the teacher for all others

The order of problems can be an important feature of the learning:

BASEBALL: importance of order of observations depends on the threshold setting for accepting hypotheses as true
CEG: order of problems is important when solutions are saved
COUNT: order of problems is very important

In all cases, when there is an ordering of problems, it is the responsibility of the teacher. In COUNT, the order of problems is critical; if COUNT is over-faced with too hard a problem, it will exceed its search tolerance without finding a solution. The art in teaching COUNT is selection of a sequence of problems that challenge it enough to learn things it couldn't do before but not to ask it to make too large a leap in one problem. BASEBALL is sensitive to ordering if the hypothesis acceptance threshold is set low; early acceptance of a mistaken hypothesis can cause difficulties in subsequent interpretation and generalization.

4. General Issues in Learning

Thus, we see that learning systems in addition to requiring various submodules [Smith et al 1977] can be described along several dimensions. Some of the dimensions of this description are:

1. Presence of Teacher: From strongly taught systems such as COUNT, CEG and Winston's to minimally taught systems such as BASEBALL and Lenat's. In the minimally taught systems, the teacher is often implicitly embodied as built-in evaluation functions, focus of attention thresholds, and heuristics.

2. Richness of Experience: From systems that need a rich experience such as COUNT and CEG to those like BASEBALL and Winston's that don't. For the latter systems, experience is often implicity given to the system in the form of schemas and descriptors for the domain. The former systems can potentially deal with more diversity in their discoveries, although they might not know what to do with it, while the latter have already been focused to interpret their experiences within a given framework or model. For these latter systems, to handle more diversity, say a Roman arch, the system needs to know when to let the model worlds bifurcate.

3. Style of Learning: The styles can range from focussed to exploratory. When one knows something about the domain or general area in which the learning takes places one can be more focussed and directed;

BASEBALL knows a lot about competitive games in general and Winston's program, about the blocks world. They both have access to symbolic descriptions and frameworks like "action" entities and "must/must not" links. At the other end of the spectrum, COUNT is like a tyro just beginning to explore its world; it needs to gather lots of experience with its primitive capabilities. CEG is somewhere intermediate on the tyro-expert learner spectrum; it is able to harness its knowledge somewhat symbollically (by knowing links between examples and relations between procedures and features), but must still do a large amount of exploration.

4. Grainsize of Knowledge: There is a spectrum of knowledge grainsize ranging from atomic primitives in COUNT, to mid-size entities and relations in CEG, to larger chunks in Winston, to large frameworks in BASEBALL.

5. Conclusions

From our own and others' experience with learning systems, it is clear that examples and experiences play a critical role in learning. While the importance of examples in learning is often overlooked, the number, variety and order of examples cannot be since they so clearly influence the style and content of learning.

In addition, while it might be fine for a system to do high-level processing when it knows something, it might be more appropriate to rely on low-level processing (e.g., trial-and-error, success-failure correlations) when it is just beginning. Perhaps, such a low-level style is the only way for the inexperienced learner and perhaps, it is a way for him to discover larger clusters of knowledge.

References

Hawkins, D., "The View from Below". For the Learning of Mathematics. Volume 1, No. 2, FLM Publishing Association, Quebec, Canada, November 1980.

Lakatos, I., Proofs and Refutations. Cambridge University Press, London, 1976.

Lenat, D. B., Automatic Theory Formation in Mathematics, Proceedings Fifth IJCAI, 1977.

Rissland, E. L., Example Generation. Proceedings Third National Conference of the Canadian Society for Computational Studies of Intelligence, Victoria, B.C., May 1980.

_____,Understanding Understanding Mathematics. Cognitive Science, Vol. 2, No. 4, 1978.

Rissland, E. L., and E. M. Soloway, Overview of an Example Generation System. Proceedings First National Conference on Artificial Intelligence, Stanford, August 1980.

_____, Constrained Example Generation: A Testbed for Studying Issues in Learning. Submitted to Seventh IJCAI, 1981.

Selfridge, O.G., Learning to Count: How A Computer Might Do It. Bolt Beranek and Newman, Inc, 1979.

Soloway, E. M., Learning = Interpretation + Generalization: A Case Study in Knowledge-Directed Learning. COINS Technical Report 78-13, University of Massachussetts, 1978.

Smith, R.G., T. M. Mitchell, R. A. Chestek and B. G. Buchanan, "A Model for Learning Systems". Proceedings Fifth IJCAI, 1977.

Winston, P. H., "Learning Structural Descriptions from Examples" in The Psychology of Computer Vision, Winston (ed), McGraw Hill, 1975.

I-Interruption effects in backward pattern masking: The neglected role of fixation stimuli

J. Pascual-Leone, J. Johnson, D. Goodman, D. Hameluck, and L. H. Theodor
York University, Toronto

Visual backward masking occurs when presentation of a visual stimulus (the mask) impairs perception of a prior briefly-presented visual stimulus (the target). A fixation stimulus (e.g., a dot or four dots demarcating a square) directs the subject's attention (i.e., fixation) to the place where target and mask will appear.

Two parameters often manipulated in backward masking research are: energy properties of mask and target, and time lag between appearance of the target and appearance of the mask (stimulus onset asynchrony-SOA). Two sorts of explanation exist for backward masking: (1) interruption theory (e.g., Turvey, 1973; Wellige et al., 1979), which interprets backward masking as a time-dependent central phenomenon, i.e., interruption by the mask's signal of processing of the target's signal; and (2) sensory integration theory (e.g., Felsten & Wasserman, 1980), which posits that the mask impairs the perception of the target by causing a peripheral or central fusion of the two sensory signals.

We present here a series of experiments which establish the importance of a new process-analytical model of backward masking. The model predicts very strong and developmentally stable, heretofore unknown, effects that no available perceptual theory can explain. One of these effects is the role of size of the fixation stimulus in moderating the disrupting impact of the mask on recognition of the target. This moderating effect of size of fixation stimulus suggests the existence of content-free central attentional processes, which regulate the influence of the mask on target recognition. After Piaget, we call this fixation-stimulus effect a centration effect; we describe briefly below the process theory and its task-analytical model of backward masking that led to the prediction of this effect.

The Theory of Constructive Operators (TCO) (Pascual-Leone & Goodman, 1979) formulates in functional terms the psychological processes which within the organism co-determine performance. These processes are of two sorts: (1) Situation-specific processes or "soft-ware" of the organism, which the theory conceptualizes as schemes; and (2) situation-free processes or "hard-ware" of the psychological organism, which the theory conceptualizes as silent operators (i.e., resources). At least four silent operators are needed to account for cognitive performances: the L, M, I, and F operators.

The L operator is the Logical learning operator. The result of L learning is the formation of habitual structures which represent frequent patterns of co-activation and of co-application of schemes. Executives are L structures that represent goal-directed generic sequences of schemes (i.e., plans). In a given situation, the dominant executive mobilizes the silent operators to produce performance.

The M and I operators represent, together, the mental effort that the organism applies on schemes to ensure that the relevant schemes will produce performance. The I operator, I for interruption (hereafter called I-interruption), represents the active inhibition that the dominant executive can monitor and use to inhibit schemes (including other conflicting executives) irrelevant for its executive-planned performance. The M operator, M for mental, is the mental attentional energy that the dominant executive mobilizes and allocates to increase the activation of (i.e., boost) task-relevant schemes. Attentional M-energy is a limited resource, able to boost only a small number of schemes simultaneously. The cardinal number of the largest set of schemes that M can boost at any one time is called the M power, and has been shown to increase developmentally (e.g., Pascual-Leone & Goodman, 1979).

The F operator, F for subjective/cognitive Field factor, corresponds to the tendency to structure performance in congruence with the subjectively salient structural features of the situation.

The operative system constituted by the executives and by the M, I, and F operators, that the dominant executive monitors and allocates in any given situation, constitute the TCO-model of the Centration Mechanism.

Consider the general backward masking paradigm we employed: after adaptation to the fixation stimulus, one of three specified target letters is flashed for 9 msec., the fixation stimulus reappears for 91 msec., following which a line-pattern mask is presented for 50 msec. Subjects must report one of the three possible targets, guessing if necessary. The masking effect occurs to the extent that the Orienting Reflex (OR) reaction (an innate executive) elicited by the mask I-interrupts processing of the target, thus allowing the target and mask signals to fuse. Recognition of the target is possible to the extent that the target executive elicited by the task instructions (i.e., "try to recognize and report the target letter") is stronger than the OR reaction so as to I-interrupt the schemes of the mask while M-boosting the schemes of the target. The target executive's task should be facilitated greatly if the target as a whole falls topographically in a retinal area which is already being M-centrated by virtue of the fixation stimulus, but should be difficult if the discriminant features of the letter fall outside the M centration. This is so because the discrimination among schemes which are "outside" versus "inside" the M centration is primary (perhaps innate) according to the theory; and the theory prescribes that the organism considers as relevant schemes which in the course of the task appear to be inside M (M-boosted). Thus, the target executive's I-interruption of irrelevant mask processes should be much easier to achieve if no task-relevant scheme (no perceptual feature of the target) falls outside the M centration initiated by the fixation stimulus.

We modified the central backward masking paradigm by including three fixation-stimulus conditions: a point fixation stimulus (a small circled dot), a square fixation stimulus (four dots demarcating a square which in size covers fully the area where the targets appear), and a square-5 fixation stimulus (a similar four-dot pattern with a fifth dot in its centre). Blocks of trials corresponding to the fixation-stimulus conditions were presented in a within-subject design. Since the square fixation will lead the subject to M-centrate the entire area in which the target will appear and the point will lead him to M-centrate only a portion of that area, we predict that the masking effect will be stronger in the point condition. With the square-5 fixation, subjects may look at either the square dots or the central point and, thus, square-5 performance should be between the other two.

A detailed task analysis of backward masking suggests that an M power (Mp) of e+4 (e for the executive's M-boost) should suffice to handle the task easily, provided that the target executive is given sufficient processing time. Much experimental-developmental work (e.g., Pascual-Leone & Goodman, 1979) has demonstrated that Mp e+4 is not available till 9-10 years of age. Thus, we predict that 9-10 year olds should perform better than younger subjects and that older subjects will not perform better than 9-10's. This age effect should occur when sufficient target-processing time is allowed (Experiments 3A and 3C below), but may not ap-

pear when target-processing time is decreased (Experiments 1, 2 and 3B).

We also predict a main effect for cognitive style field dependence/independence. Field dependent (FD) subjects should experience stronger masking effects because of their strong field (F) factor which will lead them to fuse the target and mask, and their weak I-interruption ability (Pascual-Leone & Goodman, 1979). Field independents (FI) with their strong I-interruption ability should perform best, with field mediums (FM) performing in between. Since the cognitive style effect can manifest itself only if the subject is not hindered by insufficient Mp, we make this prediction only for adults and 11- and 12-year olds. In addition, we predict that the square-point performance disparity will be greatest for FI (or FM in children) subjects because their greater I-interruption ability will enable them to best take advantage of the facilitating square.

To test our model we conducted a series of studies.

General Method

Subjects were group-tested with measures of field dependence/independence (FDI) and selected across the full FDI continuum. Children were also tested with a group measure of M power.

The tachistoscopically-presented stimuli were prepared on white cards. Two types of fixation stimulus were prepared: the point fixation, i.e., a red dot in the centre of the card with a small green circle hand-drawn around it (.25° visual angle); and the square fixation, i.e., four black dots (.15° each) arranged in a square-shaped pattern (0.9° x 0.9°) around the centre of the card. The target stimuli (the Geotype black letters A, T, and U 0.9° high x 0.7° wide) fell within the fixation area defined by the square pattern. The mask stimulus, a variant of Turvey's (1973) pattern mask, was made of black lines the same width as the target letter lines (0.2° wide).

The subject's task was to indicate which of the three specified target letters had been presented on a given trial (guessing if necessary). On each trial the subject attended to a fixation stimulus, then a target letter was presented followed by the fixation again (or blank field for adults) and then the pattern mask. The subject responded after the mask went off. The experiments were run in blocks of fixation conditions with 18 trials per block. Within a block each target letter was presented six times in random order.

Experiment 1

Method. Three age samples of 20 children each were selected; mean ages were 8;0, 10;0, and 11;10. A three-field tachistoscope was used; the fields were set to equal luminance by visual estimation. On each trial a fixation stimulus was followed by a target letter exposed for 9 msec., then the fixation for 71 msec., then the mask for 50 msec. (80 msec. SOA). Each condition had three introductory trials using 100x, 10x, and 1x the standard target duration. There were four blocks of 18 trials each. The fixed order of blocks was: point, square, point, square.

Results. An age x order x fixation ANOVA on the number of correct responses yielded a significant main effect for fixation ($F(1,57)$ 85.87, $p < .01$), with better performance in the square condition ($M=15.1$) than in the point condition ($M=11.2$). A significant order x fixation interaction ($F(1,57)$ 5.67, $p < .05$) reflected an increase in performance from the first to the second point block. Each square block showed significantly higher performance than either point block.

To illustrate the point versus square performance disparity, Figure 1 shows mean performance in the four fixation blocks for Experiments 1 and 2 (to be discussed below). The same general pattern shown here was found in all experiments.

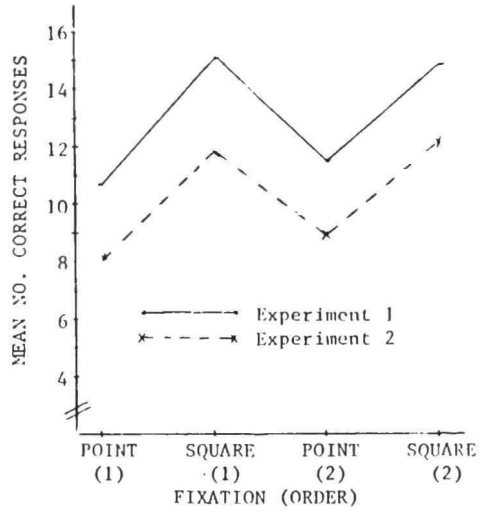


Figure 1. Mean performance across fixation blocks in Experiments 1 and 2.

Experiment 2

Method. Three age samples were selected; mean ages (and n's) were 7;4 (15), 9;4 (18), and 11;4 (15). Procedures were identical to Experiment 1.

Results. An age x order x fixation ANOVA yielded a main effect for fixation ($F(1,45)=74.11$, $p < .01$); again performance with the square ($M=12.1$) was superior to that with the point ($M=8.5$). See Figure 1.

Experiment 3

Method. Seven I.Q.-normal age-group samples were selected; mean ages (and n's) were 6;3 (21), 7;4 (24), 8;4 (24), 9;6 (20), 10;4 (25), 11;4 (23), and 12;6 (24). A three-field tachistoscope was used; its fields were set at a luminance of 6.5 cd/m^2. Introduction and fixation-block order were the same as in Experiments 1 and 2. The 6- to 8-year age samples were initially tested (treatment condition 3A) as follows: fixation stimulus, target for 9 msec., fixation for 91 msec., then mask for 50 msec. (SOA=100 msec.). The 9- to 12-year age samples were initially tested (treatment condition 3B) as follows: fixation, target for 7 msec., fixation for 93 msec., then mask for 50 msec. (SOA=100 msec). Weeks after initial testing, the samples of 8 to 12 years were re-tested (treatment condition 3C), under the conditions 3A given above.

Results. The three treatment conditions are presented successively.

3A (9 msec. target). An age x order x fixation ANOVA yielded a significant main effect for fixation ($F(1,66) = 66.2$, $p < .01$); performance with the square ($M=9.4$) was better than that with the point ($M=6.9$). Other significant effects were for age ($F(2,66)$ 3.7, $p < .05$), with 6-year-olds doing less well than older subjects; and age x order ($F(2,66)$ 3.7, $p < .05$), with older subjects generally improving with practice (i.e., over blocks) more than 6-year-olds.

3B (7 msec. target). The ANOVA yielded a

significant main effect for fixation ($F(1,88) = 40.5$, $p < .01$); performance was better with the square (M=8.8) than with the point (M=7.0). There was an order x fixation interaction ($F(1,88) = 9.5$, $p < .01$), with performance decreasing from first to second square presentation; still, each square block was significantly easier than either point block.

3C (9 msec. target, re-test). The ANOVA yielded again a significant main effect for fixation ($F(1, 111) = 145.8$, $p < .01$); square performance (M=11.4) was better than point performance (M=8.7). There were significant effects for age ($F(4,111) = 4.08$, $p < .01$), order x fixation ($F(1,111) = 5.93$, $p < .05$), and age x order x fixation ($F(4,111) = 2.98$, $p < .05$). Overall performance peaked at 9 and 10 years followed by a dropping-off at 11 and 12 years. The order x fixation interaction rested on increased performance in the point condition from first to second presentation. We will not discuss the three-way interaction, but must emphasize that for all ages except 11 years, each square block was significantly easier than either point block.

To test style predictions, 11- and 12-year-olds in 3C were divided into FD, FM, and FI groups using scores on a group version of the Children's Embedded Figures Test (CEFT - Karp & Konstadt, 1963). An age x style x order x fixation ANOVA yielded (among other significant effects) a main effect for style ($F(2,41) = 11.49$, $p < .01$), with FI's and FM's performing better than FD's; and a style x fixation interaction ($F(2,41) = 4.17$, $p < .05$), with disparity between FM and FD performance greater with the square than with the point. Pearson r's between CEFT and mean performance were $r(21) = .41$, $p = .05$ for 11's and $r(22) = .60$, $p = .002$ for 12's.

Experiment 4

Method. Thirty university undergraduates were selected to form FD, FM, and FI cognitive style groups (n = 10 in each), using scores on the Group Embedded Figures Test (GEFT - Oltman, et al., 1971). A four-field tachistoscope was used. Fixation was followed by a target stimulus of 7 msec. duration, then a 0.1 cd/m^2 full-field illumination for 93 msec., and then the pattern mask for 50 msec. (SOA = 100 msec.). All fields, except for the interstimulus interval, were set at a luminance of 6.5 cd/m^2. In addition to the square and point fixations employed in the previous experiments, a square-5 fixation was used. The square-5 was like the square, but with an additional dot (.15°) in the centre of the visual field (i.e., in the centre of the square). Again each fixation condition was run within-subjects in two blocks of 18 trials each. A training block of 18 trials preceded the experimental fixation conditions; for the training block the fixation stimulus was an "x" exactly half the size of the square. The order of fixation blocks was: training, square, square-5, point, square, square-5, point.

Results. A style x order x fixation ANOVA yielded significant main effects for style ($F(2,27) = 7.77$, $p < .01$) and fixation ($F(2,54) = 18.90$, $p < .01$). Recognition increased with degree of field independence, and decreased from the square (M=8.5) to the square-5 (M=7.0) to the point (M=6.0) conditions. The Pearson r(28) between GEFT and mean performance was .57, $p < .001$. There was a significant style x fixation interaction ($F(4,54) = 2.89$, $p < .05$) which is plotted in Figure 2. This interaction shows

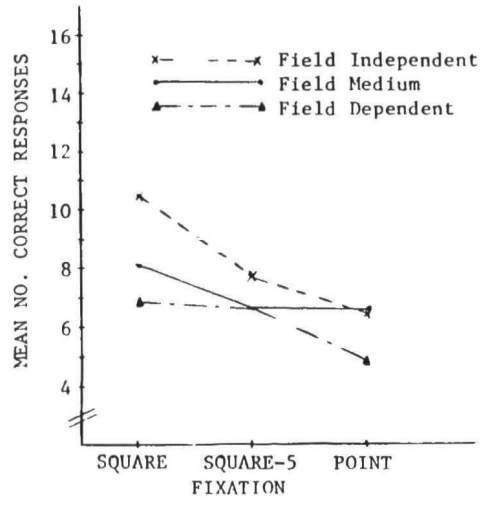


Figure 2. Mean performance by cognitive style and fixation in Experiment 4.

that the facilitating effect of the square fixation relative to the square-5 and point is greatest for FI subjects, and the misleading effect of the point fixation is greatest for FD subjects. Note that as the theory predicts, FM subjects perform between FI and FD subjects, and performance with the square-5 falls between that with the square and the point fixations.

References

Felsten, G., and Wasserman, G. S. Visual masking: Mechanisms and theories. Psychological Bulletin, 1980, 88, 329-353.

Hellige, J. B., Walsh, D. A., Lawrence, V. W., and Prasse, M. Figural relationship effects and mechanisms of visual masking. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 88-100.

Karp, S. A., and Konstadt, N. L. Manual for the Children's Embedded Figures Test. Cognitive Tests, P.O. Box 4, Vanderveer Station, Brooklyn, New York, 1963.

Oltman, P. K., Raskin, E., and Witkin, H. A. Group Embedded Figures Test. Palo Alto, Calif.: Consulting Psychologists Press, 1971.

Pascual-Leone, J., and Goodman, D. Intelligence and experience: A neoPiagetian approach. Instructional Science, 1979, 8, 301-367.

Turvey, M. T. On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with pattern stimuli. Psychological Review, 1973, 80, 1-52.

COGNITIVE LOAD AFFECTS EARLY BRAIN POTENTIAL INDICATORS OF PERCEPTUAL PROCESSING

Francois Richer
and
Jackson Beatty

Department of Psychology
University of California, Los Angeles
Los Angeles, California 90024

A major problem in cognitive psychology is the determination of the effects of primary tasks on concurrent processing of other information. Cerebral event-related potentials (ERPs) offer one approach to this problem by examining the influence of task-induced cognitive load on discrete brain events related to the processing of stimuli. In dual task situations, the introduction of tasks such as discrimination or problem solving has been found to attenuate various components of the ERP to stimuli in the secondary task (Defayolle, Dinand, & Gentil, 1971; Lille, Audin, & Hazemann, 1975). Systematic manipulations of processing demands were studied by Isreal, Wickens, Chesney, and Donchin (1980) who report an attenuation effect on the P300 component of the ERP with increasing display load in a tracking task, where workload was also assessed with a secondary reaction time task.

In a dual task the ERP-eliciting stimulus is generally relevant to a task and consequently has to be processed at a relatively high level which precludes the observation of the effects of cognitive load on earlier stages of stimulus processing. The present study was undertaken to observe the effects of cognitive load on task-irrelevant stimuli using pupil size as an independent measure of processing demands during problem solving (see Beatty, 1979).

ERPs were obtained to irrelevant auditory probes presented before and after the presentation of digits in a mental multiplication task. This task has been shown to produce pupillary dilations with a peak amplitude that is monotonically related to the a priori difficulty of the problems (Hess & Polt, 1964; Payne, Parry, & Harasymiw, 1968).

Method

Subjects and Material

Eleven college level students of both sexes were used in the experiment. The EEG and vertical electro-oculogram (EOG) were recorded with Ag-AgCl electrodes placed at Fz, Cz, Pz, F3, F4, P3, P4, referred to linked ear lobes and for the EOG from electrodes placed above and below the right eye. All signals were amplified by Grass series 8 amplifiers with a 0.1 to 70 Hz bandpass. Pupillary diameter was measured using a Whittaker 1050S video pupillometer.

Data collection and stimulation was controlled by a PDP-11/34 computer. All signals were sampled at 200 Hz for one second starting 50 msecs prior to stimulation and were digitized on-line at 12 bits. Digitized single-trial data were stored on magnetic tape for later analysis.

The auditory prompts and digits were presented through digitized speech and had a duration of 0.6 secs. The irrelevant stimuli consisted of 1000 Hz tones presented at an intensity of approximately 56 dB SL with a duration of 20 msecs. The tones were generated by a Hewlett-Packard audio oscillator (model 200 AB) and delivered through a loudspeaker 1.5 m in front of the subject. Subjects had access to a keyboard to start the trials and enter their answers on every trial.

Procedure

After the electrode placement, the subject was told that he would be required to solve 30 easy and 30 difficult multiplication problems in a random sequence. Problems were called easy when both numbers were between 2 and 9 and difficult when one number was between 3 and 9 and the other was between 13 and 19.

The procedure for each trial, as described to the subject was as follows. The subject heard the word "Ready" through the loudspeaker at which time he was to fixate a dot in front of him and depress a key to start the trial. He was then to attend to the auditory presentation of the two digits separated by a one-second pause and to ignore all other stimuli. After the second digit the subject was to silently multiply the two digits until he heard the word "Respond", at which time he could stop fixating and was to enter the correct answer on the keyboard.

A tone was presented at a random time between one and two seconds after the subject's initiation of the trial and also between one and two seconds after the presentation of the last digit. The interval between the end of the last digit and the response cue varied between 3.5 and 4.5 secs.

Results

For each subject, the individual EEG epochs were edited for artifacts using an RMS voltage criterion rejecting an average of 10% of the trials. The epochs were then averaged separately for each of the four conditions defined by the two recording periods and the two difficulty levels. Averaged pupillary responses were also obtained using the same trials. Three of the initial eleven subjects had to be excluded from the analysis because their EEG records contained a large amount of ocular artifacts.

Two main components could be identified in all the averaged waveforms by visual inspection: a negative one having a latency ranging from 80 to 130 msecs (N100) and a positive one with a latency range of 170 to 220 msecs (P200).

Group averages were obtained (see Figure 1) for the four conditions and difference waveforms were computed between ERPs in baseline and problem solving periods. Slow potential shifts could not be discerned in any of the difference waveforms.

The peak amplitude of the components relative to an average pre-stimulus baseline was obtained using the maximum value within the latency ranges described above. At Fz, N100 amplitude was found to be generally lower during problem solving periods than during baseline periods ($F1,7)=11.7$, p .01).

At Cz the mean baseline amplitude was slightly higher for the difficult condition compared to the easy condition but a much lower amplitude was observed in the difficult condition during the task compared to the other three conditions. Significant effects were obtained for the main effect of task presence ($F(1,7)=9.36$, $p < .05$) and for the interaction between task presence and difficulty level ($F(1,7)=5.66$, $p < .05$). An analysis of the simple effects in this interaction revealed that task presence produced a significant amplitude decrement on N100 only in the difficult condition ($F(1,7)=10.8$, $p < .025$) and not in the easy condition ($F < 1$).

No significant effects were observed on the N100 component at any other derivation and no effects were found on the amplitude of P200 at any electrode site.

Average pupil size at stimulus onset was also analyzed using ANOVA and significant effects were obtained for the level of difficulty ($F(1,7)=6.17$, $p < .05$) and for the interaction between task presence and difficulty level ($F(1,7)=38.27$, $p < .01$). Analyses of simple effects determined that, as with N100 amplitide, the effect of the task on pupil diameter was only significant in the difficult condition ($F(1,7)=8.4$, $p < .025$). Figure 2 contrasts the effects obtained on pupil size and on N100 amplitude at Cz.

In order to examine the relation between the effects observed on the pupil and N100 measures across subjects rank correlations were obtained between the changes in pupil size and in N100 amplitude from baseline to task periods across subjects for both the easy and difficult conditions. A significant negative correlation was obtained in the easy condition ($p=-0.77$, $p < .05$) but not in the difficult condition ($p=0.22$, N.S.).

Discussion

The present data suggest that the amplitude of the N100 component evoked by irrelevant stimuli can be affected by cognitive load as indexed by pupillary measures.

The fact that no N100 attenuation could be observed in the easy trials argues against an interpretation of the results in terms of a general activation effect due to the performance of a task. The same argument can be used to refute significant effects due to the long recovery period of N100 with multiple stimulation, and significant contamination due to response preparation or CNV-like potentials. Multiplication problems involving single digit pairs have been shown to produce identifiable pupillary dilations in other studies (e.g., Hess & Polt, 1964) but in the present experiment the second probe was presented rather late after the last digit and, on most trials, probably occurred after the resolution of the pupillary response.

The absence of slow potential shifts in the global difference waveforms is not sufficient to conclude that the N100 attenuation is not due to variations in overlapping potential shifts. It is clear however, that the influence of the task on the ERP takes place at an early stage. Indeed, the N100 component has been shown to be closely associated with processes related to the early perceptual strength of stimuli indexed by sensitivty measures in signal detection situations (e.g., Squires, Hillyard, & Lindsay, 1973) and intensity parameters in multichannel selective listening situations (Schwent, Hillyard, & Galambos, 1976). The finding that the cognitive load induced by problem solving can affect CNS activity at latencies similar to those of processes related to the strength of physiological signals can have an important impact on conceptualizations of information processing. These suggestive results will have to be complemented by more direct observations to establish the nature of this possible influence of problem solving on early stages of perceptual processing.

References

Beatty, J. Pupillometric methods of workload evaluation: Present status and future possibilities. In B.O. Hartman, & R.E. McKenzie (Eds.), Survey of methods to assess workload. London: AGARD (AG No. 246), 1979.

Defayolle, M., Dinand, J.P., & Gentil, M.T. Averaged evoked potentials in relation to attitude, mental load and intelligence. In W.T. Singleton, J.G. Fox, & D. Whitfield (Eds.), Measurement of man at work. London: Taylor and Francis, 1971.

Hess, E.H., & Polt, J.H. Pupil size in relation to mental activity during simple problem solving. Science, 1964, 143, 1190-1192.

Isreal, J.B., Wickens, C.D., Chesney, G.L., & Donchin, E. The event-related brain potential as an index of display-monitoring workload. Human Factors, 1980, 22, 211-224.

Lille, F., Audin, G., & Hazemann, P. Effects of time and tasks upon auditory and somatosensory evoked potentials in man. Electroencephalography & Clinical Neurophysiology, 1975, 39, 239-246.

Payne, D.T., Parry, M.E., & Harasymiw, S.J. Percentage pupillary dilation as a measure of item difficulty. Perception & Psychophysics, 1968, 4, 139-143.

Schwent, V.L., Hillyard, S.A., & Galambos, R. Selective attention and the auditory vertex potential. II.Effects of signal intensity and masking noise. Electroencephalography & Clinical Neurophysiology, 1976b, 40, 615-622.

Squires, K.C., Hillyard, S.A., & Lindsay, P.H. Vertex potentials evoked during auditory signal detection: Relation to decision criteria. Perception & Psychophysics, 1973, 14, 265-272.

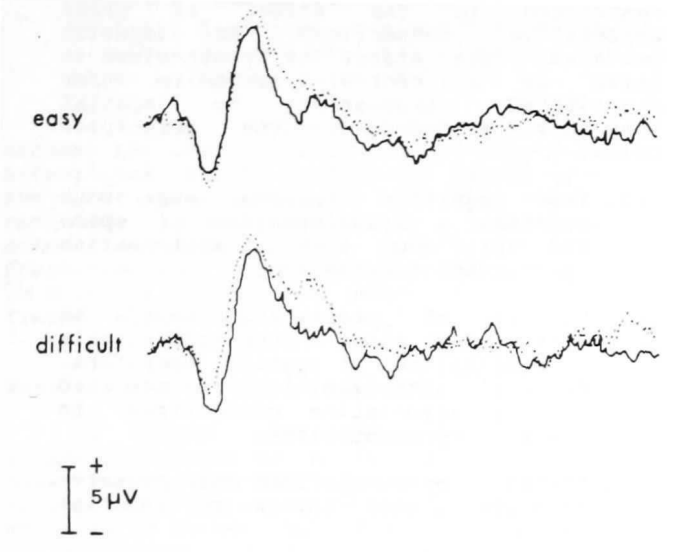


Figure 1. Group average waveforms for the two experimental conditions obtained to stimuli during baseline (solid) and problem solving (dotted) periods.

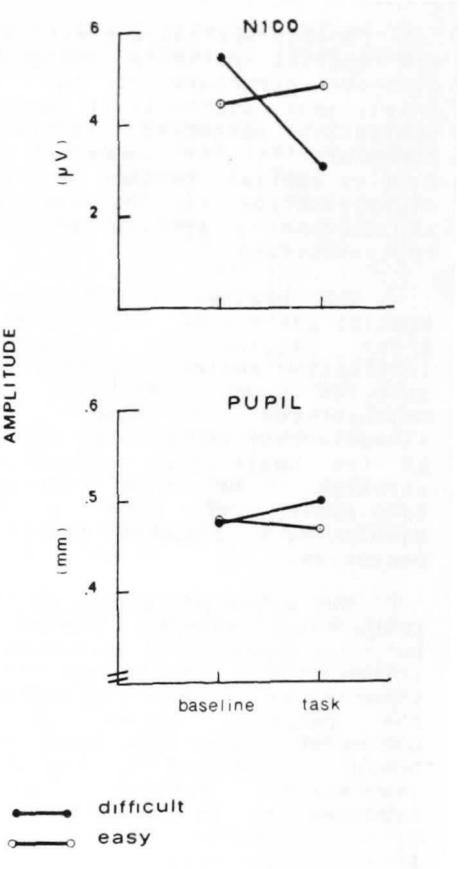


Figure 2. Mean amplitudes of N100 and mean pupil diameters in the four conditions.

THE SPATIAL REPRESENTATION AND PROCESSING OF INFORMATION IN COGNITION[1]

by
Gary W. Strong and Bruce A. Whitehead[2]

Objects spatial patterns within them, and spatial patterns among them form the concrete structure of our world, and intelligent action in it depends upon such spatially patterned information [1]. Therefore, it is important to define the role of spatial pattern in the internal representation of the environment and in the processing applied to this internal representation.

The premise of our model is that the spatial pattern of real-world information shows regularities independent of its information content. Extensive analysis of stimulus properties [2-5] have indeed demonstrated such regularities (invariances) and pointed out their value as the basis of an information processing strategy. We have investigated the feasibility of such a strategy by simulating a proposed cognitive operator mechanism.

Our model postulates an internal space image which represents the spatial pattern, but not content, of environmental stimulus information. (The properties of this internal space image correspond to those of the parietal cortex in the brain.) Cognitive operations upon this internal space correspond to transformations in real-world spatial pattern. As hypothesized in the information processing principles below, the cognitive operator serves a dual role: It represents a generic real-world action which transforms the spatial pattern of incoming information, and it carries out the corresponding cognitive operation on the spatial pattern of the internal space image. In this latter role, the cognitive operator tracks or predicts the spatial transformation of the environmental stimulus which will result from the real-world action.

Our model is highly simplified with respect to brain structure and limited to a specific problem domain: the movement of attention over the visual field. The enhancement effect in parietal cortex has been put forth as a substrate for this cognitive operation. While a cognitive model such as ours cannot show detailed neural mechanisms, it can be constrained to be neurally feasible and to be consistent with the neural data, as discussed under "Physiological Constraints" below.

[1]This research supported in part by National Science Foundation Grant No. IST-8011617.

[2]Systems Science Institute, University of Louisville, Louisville KY 40292

Hypothesized Information Processing Principles

The simulation model implements hypothesized principles by which neural interconnection structure represents spatial pattern information about the environment. Content representations are modeled only as they participate in the control of the spatial processing system. More precisely, distributed processing of spatial information is influenced by associations between the spatial and the figural representation systems. The role of figural representations in the model is restricted to the effect of their associative connections to spatial patterns. This effect is hypothesized to serve as a control mechanism for distributed processing by spatial operators, according to the principles below:

(a). The cognitive operator must both represent a transformation of space and carry out such a transformation upon object information.

(b). Control of processing must be distributed among content-addressable representations of spatial operators. Content addressability is achieved through associative connections to figural representations.

(c). Content representations must preserve identity across change of spatial locus.

(d). The cognitive operator must be able to carry out spatial transformations independently of object content.

Principles (a) and (b) have been developed in previous work [6]. Efficient access to the information in a representation requires content-addressability. However, this content addressability can be achieved by competition among the representations themselves acting as independent, parallel processors. This avoids the need for centralized control of the distributed processing network.

Physiological Constraints

The enhancement effect in parietal cortex is a neurophysiological correlate of the spatial locus of psychological attention over the visual field [7-9]. No mechanism has previously been put forth which (i) carries out the spatial transformation (Principle a above), (ii) produces its neurophysiological correlate (the parietal enhancement effect), and thus (iii) moves the psychological locus of attention. A model based upon parietal cortex implements our hypothesized mechanism. The model is designed to be consistent with neurophysiological data without attempting to be anatomically complete.

Parietal cell recording studies [7-8] demonstrate internal representations where the spatial layout in the neural tissue represents that in the environment. The anatomical coordinate system is in this sense an image of the surrounding

environmental space. In the present model, it is termed the INTERNAL SPACE IMAGE.[3] Enhancement in the response of parietal cells to a peripheral stimulus has been shown to

(e). Correspond to the environmental locus of the target stimulus [7-8].

(f). Correspond to a psychological locus of attention (Jonides' [10] "mind's eye") directed toward the stimulus [7-8].

(g). Precede not only a saccadic eye movement to the stimulus, but also a reaching hand movement to it [8], or even the directing of attention to the stimulus without making eye, hand, or other movements to it [8-9].

In sum, while neural enhancement encodes the spatial locus of an impending attentional shift, it does not depend upon the motor characteristics of the shift. Parietal enhancement is therefore associated with a more general spatial transformation than the enhancement effect in superior colliculus which is rigidly tied to eye movement shifts [9].

Resulting Model Specifications

It would seem that the parietal enhancement mentioned above should help in some way to produce a more functional percept by emphasizing retinal information which is to become the target of a movement or attentional shift. However, in the usual case where the peripheral stimulus is quickly brought to the fovea, the function of enhancement at the original peripheral neural locus is unclear. The simulation model seeks to clarify the function of this peripheral enhancement as follows:

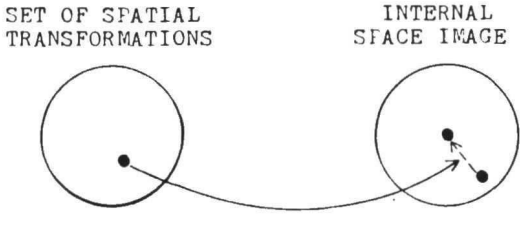


FIGURE 1: THE REPRESENTATION OF A TRANSFORMATION

In a system which represents transformations, representation of the magnitude and direction of the shift is essential. This information is defined by the enhanced peripheral locus (point p in Figure 1) in combination with the fovea (point q in Figure 1). Therefore, the function of such peripheral enhancement may be to encode such a vector in the parietal representation of space. The peripheral enhancement would then indicate the activation of the encoded vector.

The Hypothesized Information Processing Principles (a-d) and Physiological Constraints (e-g) above imply these model specifications:

(1). A fundamental transformation mechanism exists as shown in Figure 1 (discussed above).

(2). Figural content of input is separated from its spatial locus information (Constraints c and d). Only the spatial locus of input is passed to the internal space image (Constraint e). Therefore, figural content of input can affect the internal space image only in the selection of spatial transformations.

(3). Information representing spatial locus in the environment is coded by activity at corresponding loci in the internal space image (Constraint e).

(4). The figural identity of active loci in the internal space image must be represented elsewhere in the system (Constraints c and d).

(5). Since control of processing must be distributed among content-addressable representations (Constraint b), control must originate outside the internal space image.

(6). Therefore, the spatial transformations which act upon the internal space image must be selected by their associations to content representations (Constraint a).

By decomposing stimulus information into its spatial and figural components (Principle d leading to Model Specification 2), each component is independently generalizable. A new stimulus can therefore be separately matched with (i) spatial representations learned in a different figural context and (ii) figural representations learned in a different spatial context.

The remaining point to be addressed in the model is the control of distributed processing (Principle b). More specifically, only one of the operators in the spatial transformation set will be appropriate for a given stimulus. Selection of this operator must depend upon both spatial and figural properties of the stimulus. This selection process therefore requires reconvergence of the spatial and figural components of stimulus information. Appropriate spatial operators are addressed via convergent association with active figural representations. This associative content-addressing serves as the control mechanism for the distributed processing network (Model Specifications 5 and 6).

[3] "Image" is used here in the sense of a mathematical function from external to internal coordinate systems.

The Model Structure

As described above, the model operates by initial decomposition of incoming stimulus information into generic spatial and figural components and subsequent association of these components. A model structure which illustrates this and which also satisfies Model Specifications 1-6 is shown in Figure 2.

Perceptual input arrives at the circle labelled LOCUS-DEPENDENT FEATURES. Its feature content information is picked up by LOCUS-INDEPENDENT FEATURES, while the spatial locus of each input feature is simply mapped one-to-one onto a corresponding "hotspot" within the INTERNAL SPACE IMAGE. The locus of the "hotspot" in the INTERNAL SPACE IMAGE therefore represents the spatial position of the feature, but does not carry any content information. However, each "hotspot" activates a set of affordances (movement possibilities) by virtue of the fact that one or more LOCUS-DEPENDENT OPERATORS in the spatial transformation set may originate at the locus occupied by the "hotspot". Each of these LOCUS-DEPENDENT OPERATORS represents a specific, local transformation in the INTERNAL SPACE IMAGE. At any one time, numerous, conflicting possible transformations may be suggested by the current set of "hotspots" in the INTERNAL SPACE IMAGE.

The selection of an appropriate, consistent set of transformations for the model to carry out requires the convergence of activation from two different pathways: (1) from active LOCUS-INDEPENDENT FEATURES, through the (learned) ASSOCIATION NETWORK, into LOCUS-INDEPENDENT OPERATORS; and (2) from the set of possible transformations (affordances as defined above) given by the current analysis of the INTERNAL SPACE IMAGE, into LOCUS-INDEPENDENT OPERATORS. Thus activation from two pathways converges into the set of LOCUS-INDEPENDENT OPERATORS. Any such operator receiving simultaneous activation from both pathways represents a transformation that is suitable to perform, given both the spatial and content limitations of the environment. At this point, the operator selected by convergent activation may in turn activate its associated LOCUS-DEPENDENT OPERATORS, leading to actual behavior as well as to a prediction of the next internal space layout.

This model structure has been implemented and tested in a computer program. The model exhibits not only recognition of spatially-patterned objects but also fill-in of missing features in appropriate places, suggesting the successful encoding of pattern information within objects. Further testing is underway to examine its ability to form generalizations from perceptual input. We plan to adapt the model to a game, such as chess, which might simulate (in a serial program) the ability of neural circuitry to efficiently process spatially-encoded information in parallel.

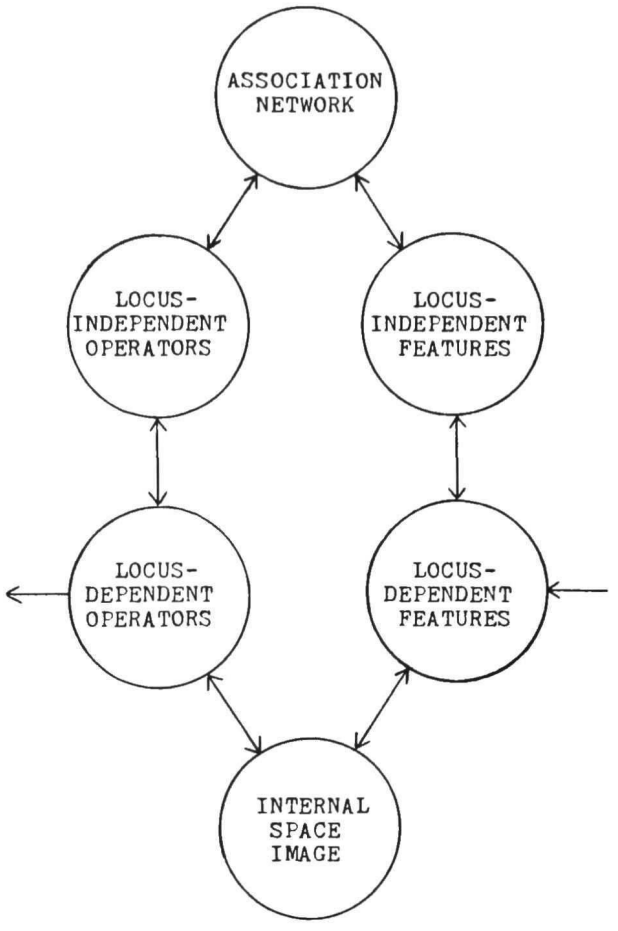


FIGURE 2: THE MODEL STRUCTURE

The important features of the model are that it encodes spatial pattern within a non-spatial association network, thereby eliminating the necessity of storing a "picture". Instead a network of stored feature-movement relations is sufficient to reconstruct the picture whenever necessary. Thus the model deals with the relational aspect of information as well as the bit-content aspect. Each feature-movement relation can be used not only to reconstruct the image of the perceived object but also to evoke behavior appropriate to the object. Finally, the model structure described can be translated into other domains requiring goal-directed information processing, such as in cultural systems [11].

REFERENCES

1. Posner, M.I., Nissen, M.J., and Ogden, W.C. Attended and unattended processing modes: The role of set for spatial location. In H.L. Pick and E. Saltzman (Eds.), Modes of Perceiving and Processing Information, Hillsdale, N.J.: Earlbaum, 1978.

2. Gibson, J.J. The Perception of the Visual World. Boston: Houghton Mifflin, 1950.

3. Gibson, J.J. An Ecological Approach to Visual Perception. Boston: Houghton Mifflin, 1979.

4. Lee, D.N. Visual information during locomotion. In R.B. MacLeod & H.L. Pick (Eds.), Perception: Essays in Honor of James J. Gibson. Ithaca: Cornell University Press, 1974.

5. Shaw, R. & Pittenger, J. Perceiving change. In H.L. Pick, Jr. & E. Saltzman (Eds.), Modes of Perceiving and Processing Information. Hillsdale, N.J.: Lawrence Earlbaum, 1978.

6. Whitehead, B.A. A Neural Model of Human Pattern Recognition. New York: Garland, in press.

7. Yin, T.C. and Mountcastle, V.B. Visual input to the visuomotor mechanisms of the monkey's parietal lobe. Science, 1977, 197, 1381-1383.

8. Robinson, D.L., Goldberg, M.E., and Stanton, G.B. Parietal association area in the primate: sensory mechanisms and behavioral modulations. Journal of Neurophysiology, 1978, 41, 910-932.

9. Robinson, D.L., & Goldberg, M.E. The visual substrate of eye movements. In J.W. Senders, D.F. Fisher, & R.A. Monty (Eds.), Eye Movements and the Higher Psychological Functions. Hillsdale, N.J.: Earlbaum, 1978.

10. Jonides, J. Voluntary Versus Reflexive Control of the Mind's Eye's Movement. Paper presented to the Psychonomic Society, November 1976.

11. Strong, G.W. Information, Pattern, and Behavior: The Cognitive Biases of Four Japanese Groups. Ph.D. Dissertation, University of Michigan, Ann Arbor, 1981.