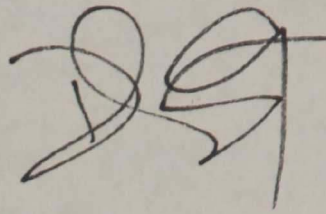


Gentner



Proceedings of the  
**Fifth Annual  
Conference  
of the  
Cognitive  
Science Society**

Sponsored By The Cluster in Cognitive Science, University of Rochester

Supported in part by a Grant from the Alfred P. Sloan Foundation

**Rochester, New York**

**May 18-20, 1983**





Proceedings of the  
**Fifth Annual  
Conference  
of the  
Cognitive  
Science Society**

Sponsored By The Cluster in Cognitive Science, University of Rochester

Supported in part by a Grant from the Alfred P. Sloan Foundation

**Rochester, New York**

**May 18-20, 1983**



## Preface

This volume contains the submitted papers which were presented at COGSCI 5, the fifth annual meeting of the Cognitive Science Society which was held at the University of Rochester during May 18 through 20, 1983.

This was the first Cognitive Science meeting in which submitted papers were reviewed for acceptance. (Only about half of them were accepted since we received about twice as many submissions as there were slots in the program for them to be presented in.) The reviewing panel consisted of James Allen, Jerome Feldman, Stephen Kosslyn, Pat Siple and Michael Tanenhaus as well as myself, and outside reviews were obtained from many others.

About half of the conference consisted of invited addresses and panel discussions, which are not reproduced here.

Organizing this meeting has been an exhilarating and exhausting experience, especially as it was done unexpectedly soon (we originally volunteered for 1984) and unexpectedly quickly (the fourth conference was in August, we in May). Most of the actual work, however, has been done by Rita Fay and Margery Lucas, to whom go our heartfelt thanks.

Patrick J. Hayes	Chairman
David Taylor	Local Arrangements
Sheldon Wagner	Banquet Organizer



# PROGRAM

## The Fifth Annual Conference of the Cognitive Science Society

May 18, 19, 20, 1983

The University of Rochester  
Rochester, New York



## Program

All conference sessions will be held in the Mathematical Sciences Building and Hutchinson Hall on the University of Rochester campus.

### Tuesday, May 17

5:00-8:00 p.m.

**Registration**, Susan B. Anthony Hall, Green Room

7:00-11:00 p.m.

**Welcoming Reception**, Eastman Quadrangle

### Wednesday, May 18

8:00 a.m.-5:00 p.m.

**Registration**, Mathematical Sciences Building, Red Lounge

9:00-10:30 a.m.

**Invited Address**, Hubbell Auditorium

"Neural Systems of Attention"

*Michael Posner*, University of Oregon

10:30-10:45 a.m.

Morning Break

10:45 a.m. Hubbell Auditorium

**Panel**: Decision-Making

Participants:

*Henry Kyburg*, Chair – University of Rochester

*Daniel Kahneman* – University of British Columbia

*Amos Tversky* – Stanford University

*Isaac Levy* – Columbia University

*Teddy Seidenfeld* – Washington University

12:30-2:00 p.m.

Lunch Break – Danforth Dining Center, Susan B. Anthony Hall

2:00-3:20 p.m.

**Paper Session #1**, Hutchinson Hall

2:00-2:20 p.m.

"Lexical Access During Sentence Comprehension: Frequency and Context Effects"

*Margery Lucas*, University of Rochester

2:20-2:40 p.m.

"Comprehending and Remembering Novel Metaphors While Reading Real Text"

*Judith Orasanu*, U.S. Army Research Institute

2:40-3:00 p.m.

"Verb Semantic Structures and On-Line Language Processing"

*Ronan G. Reilly*, St. Patrick's College

3:00-3:20 p.m.

"Lexical Access and Serial Order in Sentence Production"

*Joseph Paul Stemberger*, Carnegie-Mellon University

2:00-3:20 p.m.

**Paper Session #2**, Hubbell Auditorium

2:00-2:20 p.m.

"Computational Selection of Processing Locations in Vision"

*Roger Browse*, Queen's University

2:20-2:40 p.m.

"Topical Invariants and Apparent Motion"

*Lin Chen*, China University of Science & Technology and University of California at Irvine

2:40-3:00 p.m.

"Imagined Spatial Transformations in the Visual Discrimination of Left and Right Parts of the Human Body"

*Larry Parsons*, University of California at San Diego

3:00-3:20 p.m.

"The Structure of Visual Concepts"

*Joseph Psotka*, Army Research Institute

2:00-3:20 p.m.

**Paper Session #3**, Room #201

2:00-2:20 p.m.

"The Role of Interpersonal Games in Perlocutionary Acts"

*Gabriella Airenti*, *Bruno G. Bara*, *Marco Colombetti*, Università di Milano and Politecnico di Milano

2:20-2:40 p.m.

"Ninety Years of Mental Metaphors"

*Dedre Gentner* and *Jonathan Grudin*, Bolt Beranek and Newman, Inc.

2:40-3:00 p.m.

"The Evaluation of Cognitive Constructs Using Structural Equation Modeling"

*Morton Ann Gernsbacher* & *H. Hill Goldsmith*, University of Texas at Austin

3:00-3:20 p.m.

"Countplan: A Model for Planning Counting Procedures Based on Utilization Knowledge, and Procedural and Conceptual Competence"

*Donald A. Smith* and *James G. Greeno*, University of Pittsburgh

3:20-3:35 p.m.

Break

3:35-5:00 p.m.

**Invited Address**, Hubbell Auditorium

"Intuition, Reasoning, and Multilevel Concurrent Processing"

*David Waltz*, University of Illinois

5:00-7:30 p.m.

Dinner Break – Wilson Commons

7:30-9:30 p.m. Hubbell Auditorium

**Panel:** The Computational Metaphor

Participants:

*Patrick J. Hayes*, Chair – University of Rochester

*Zenon Pylyshyn* – University of Western Ontario

*Brian Smith* – Xerox PARC

*John Haugeland* – University of Pittsburgh

9:30-12:00 M

**Reception**, May Room, Wilson Commons

**Thursday, May 19**

8:00 a.m.-5:00 p.m.

**Registration**, Mathematical Sciences Building, Red Lounge

9:00-10:20 a.m.

**Paper Session #4**, Hutchinson Hall

9:00-9:20 a.m.

"Understanding and Solving Word Arithmetic Problems"

*Walter Kintsch* and *James G. Greeno*, University of Pittsburgh

9:20-9:40 a.m.

"Children's Mental Models of Recursive Logo Programs"

*D. Miriam Kurland* and *Roy D. Pea*, Bank Street College of Education

9:40-10:00 a.m.

"How Do Children Think About Numbers? Let Us Count the Ways"

*James A. Levine*, *Marcia J. Boruta*, *Andrea L. Petitto*, University of California at San Diego and University of Rochester

10:00-10:20 a.m.

"A Computational Description of the Three Stages of Development of Object Identity by Infants"

*George F. Luger*, *Jennifer G. Wishart* and *T.G.R. Bower*, University of New Mexico and University of Edinburgh

9:00-10:20 a.m.

**Paper Session #5**, Hubbell Auditorium

9:00-9:20 a.m.

"Reasoning About the Temporal Structure of Narrative Texts"

*Michael J. Almeida* and *Stuart C. Shapiro*, State University of New York at Buffalo



9:20-9:40 a.m.

"Event Concept Coherence in Narrative Text"

*Richard Alterman, University of Texas at Austin*

9:40-10:00 a.m.

"Perseverers, Recencies and Deferrers: New Experimental Evidence of Multiple Inference Strategies in Understanding"

*Richard H. Granger and Jennifer K. Holbrook, University of California at Irvine*

10:00-10:20 a.m.

"Stratification in Story Understanding"

*Valentina Zavarin, Wright Institute & University of California at San Francisco*

9:00-10:20 a.m.

**Paper Session #6**, Room #201

9:00-9:20 a.m.

"Multiple Models of Evaporation Processes"

*Allan Collins, Dedre Genter, Bolt Beranek and Newman, Inc.*

9:20-9:40 a.m.

"A Central Problem in Representing Human Knowledge in Artificial Systems: The Transformation of Intrinsic Into Extrinsic Representations"

*Gerhard Dirlich, Christian Freksa and Ulrich Furbach, Max-Planck-Institut für Psychiatrie, and Hochschule der Bundeswehr*

9:40-10:00 a.m.

"What Else is Wrong with Non-Monotonic Logics? Representational and Informational Shortcomings"

*Jane Terry Nutter, State University of New York at Buffalo*

10:00-10:20 a.m.

"Tacit Theories of Economics"

*William J. Salter, Bolt Beranek and Newman, Inc.*

10:20-10:40 a.m.

"Student Modeling at Strategy Learning"

*Pat Langley, Carnegie-Mellon University*

10:20-10:35 a.m.

Morning Break

10:35-11:35 a.m.

**Paper Session #7**, Hutchinson Hall

10:35-10:55 a.m.

"On Association Techniques in Neural Representation Schemes"

*John A. Barnden, Indiana University*

10:55-11:15 a.m.

"Pains and Strong Cognitivism"

*George Graham and G. Lynn Stephens, University of Alabama in Birmingham*

11:15-11:35 a.m.

"Analyzing Cooperative Computation"

*G. E. Hinton and T. J. Sejnowski, Carnegie-Mellon University and Johns Hopkins University*

10:35-11:35 a.m.

**Paper Session #8**, Hubbell Auditorium

10:35-10:55 a.m.

"Long Term Recency Effects in Recalling Previous Answers"

*M. W. Lansdale, ITT Europe*

10:55-11:15 a.m.

"The Role of Grid Schemata in Memory for Large-Scale Environments"

*Ian Moar, Nancy J. Hamer and Betsy A. Woods, Bucknell University*

11:15-11:35 a.m.

"The Roles of Interference and Inference in the Retrieval of Autobiographical Memories"

*Brian J. Reiser and John B. Black, Yale University*

10:35-11:35 a.m.

**Paper Session #9**, Room #201

10:35-10:55 a.m.

"Thematic Relations Between Episodes"

*Colleen M. Seifert and John B. Black, Yale University*

10:55-11:15 a.m.

"Reading a Program is Like Reading a Story (Well, Almost)"

*Elliot Soloway, Kate Ehrlich, Eric Gold, Yale University*

11:15-11:35 a.m.

"Cerebral Lateralization, Preferred Cognitive Mode and Reading Achievement in American Indian Bilingual Children"

*Fairlee E. Winfield, Northern Arizona University*

11:35-12:35

Business Meeting of the Cognitive Science Society

Hubbell Auditorium

12:35-2:00 p.m.

Lunch Break – Danforth Dining Center, Susan B. Anthony Hall

2:00-3:30 p.m.

**Invited Address**, Hubbell Auditorium

"A Computational View on the Pragmatics of Noun Phrases"

*Candace Sidner, Bolt, Beranek and Newman, Inc.*

3:30-3:45 p.m.  
Break

3:45 p.m. Hubbell Auditorium  
**Panel:** Cognitive Development

Participants:  
*Sheldon Wagner*, Chair – University of Rochester  
*Leslie Cohen*, University of Texas  
*Susan Carey*, Massachusetts Institute of Technology  
*Frank Keil*, Cornell University

7:00 p.m.  
**Banquet** – Faculty Club  
**Guest Speaker:** *Jerome Lettvin*, Department of Biology, Massachusetts Institute of Technology

## Friday, May 20

8:00-12:00 N  
**Registration**, Mathematical Sciences Building, Red Lounge

9:00-10:30 a.m.  
**Invited Address**, Hubbell Auditorium  
“Lexical Codes and Language Comprehension”  
*Michael Tanenhaus*, Wayne State University

10:30-10:45 a.m.  
Morning Break

10:45 a.m. Hubbell Auditorium  
**Panel:** Expert Systems

Participants:  
*John McDermott*, Chair – Carnegie-Mellon University  
*Ted Shortliffe* – Stanford University Medical Center  
*Ben Kuipers* – Tufts University  
*Allan Lesgold* – University of Pittsburgh

12:30-2:00 p.m.  
Lunch Break – Danforth Dining Center, Susan B. Anthony Hall

2:00 p.m. Hubbell Auditorium  
**Panel:** Society for Cognition and Brain Theory

2:00-4:30 p.m. **Seminar:** Neuroscience for Cognitive Science, Hubbell Auditorium (Co-sponsored by Society for Cognition and Brain Theory)

*Theodore Melnechuk*, Research Associate at Western Behavioral Sciences Institute, La Jolla, CA and Visiting lecturer, Dept. of Neurosciences, School of Medicine, University of California at San Diego, Chair

2:00-2:05:	Introduction
2:05-2:35 p.m.	“The New Neurobiology” <i>John R. Sladek Jr.</i> , University of Rochester School of Medicine
2:35-2:45 p.m.	Questions and Answers
2:45-3:15 p.m.	“A Cognitive Interpretation of a Neurobiological Experiment” <i>David Zipser</i> , University of California at San Diego
3:15-3:25 p.m.	Questions and Answers
3:25-3:55 p.m.	“Parallel Processing in Neural Modeling” <i>Terrence J. Sejnowski</i> , Johns Hopkins University
3:55-4:05:	Questions and Answers
4:05-4:30:	Panel Discussion <i>Theodore Melnechuk</i> <i>John R. Sladek Jr.</i> <i>David Zipser</i> <i>Terrence J. Sejnowski</i> <i>Jerome A. Feldman</i> , University of Rochester

## Table of Contents

### Paper Session #1

"Lexical Access During Sentence Comprehension: Frequency and Context Effects"  
Margery Lucas, University of Rochester

"Comprehending and Remembering Novel Metaphors While Reading Real Text"  
Judith Orasanu, U.S. Army Research Institute

"Verb Semantic Structures and On-Line Language Processing"  
Ronan G. Reilly, St. Patrick's College

"Lexical Access and Serial Order in Sentence Production"  
Joseph Paul Stemberger, Carnegie-Mellon University

### Paper Session #2

"Computational Selection of Processing Locations in Vision"  
Roger Browse, Queen's University

"Topical Invariants and Apparent Motion"  
Lin Chen, China University of Science & Technology and University of California at Irvine

"Imagined Spatial Transformations in the Visual Discrimination of Left and Right Parts of the Human Body"  
Larry Parsons, University of California at San Diego

"The Structure of Visual Concepts"  
Joseph Psotka, Army Research Institute

### Paper Session #3

"The Role of Interpersonal Games in Perlocutionary Acts"  
Gabriella Airenti, Bruno G. Bara, Marco Colombetti, Universita di Milano and Politecnico di Milano

"Ninety Years of Mental Metaphors"  
Dedre Gentner and Jonathan Grudin, Bolt Beranek and Newman, Inc.

"The Evaluation of Cognitive Constructs Using Structural Equation Modeling"  
Morton Ann Gernsbacher & H. Hill Goldsmith, University of Texas at Austin

"Countplan: A Model for Planning Counting Procedures Based on Utilization Knowledge, and Procedural and Conceptual Competence"  
Donald A. Smith and James G. Greeno, University of Pittsburgh



#### Paper Session #4

"Understanding and Solving Word Arithmetic Problems"

Walter Kintsch and James G. Greeno, University of Pittsburgh

"Children's Mental Models of Recursive Logo Programs"

D. Miriam Kurland and Roy D. Pea, Bank Street College of Education

"How Do Children Think About Numbers? Let Us Count the Ways"

James A. Levine, Marcia J. Boruta, Andrea L. Petitto, University of California at San Diego and University of Rochester

"A Computational Description of the Three Stages of Development of Object Identity by Infants"

George F. Luger, Jennifer G. Wishart and T.G.R. Bower, University of New Mexico and University of Edinburgh

#### Paper Session #5

"Reasoning About the Temporal Structure of Narrative Texts"

Michael J. Almeida and Stuart C. Shapiro, State University of New York at Buffalo

"Event Concept Coherence in Narrative Text"

Richard Alterman, University of Texas at Austin

"Perseverers, Recencies and Deferrers: New Experimental Evidence of Multiple Inference Strategies in Understanding"

Richard H. Granger and Jennifer K. Holbrook, University of California at Irvine

"Stratification in Story Understanding"

Valentina Zavarin, Wright Institute & University of California at San Francisco

#### Paper Session #6

"Multiple Models of Evaporation Processes"

Allan Collins, Dedre Gentner, Bolt Beranek and Newman, Inc.

"A Central Problem in Representing Human Knowledge in Artificial Systems: The Transformation of Intrinsic Into Extrinsic Representations"

Gerhard Dirlich, Christian Freksa and Ulrich Furbach, Max-Planck-Institut für Psychiatrie and Hochschule der Bundeswehr

"What Else is Wrong with Non-Monotonic Logics? Representational and Informational Shortcomings"

Jane Terry Nutter, State University of New York at Buffalo

"Tacit Theories of Economics"

William J. Salter, Bolt Beranek and Newman, Inc.

"Student Modeling at Strategy Learning"

Pat Langley, Carnegie-Mellon University





Paper Session #7

"On Association Techniques in Neural Representation Schemes"  
John A. Barnden, Indiana University

"Pains and Strong Cognitivism"  
George Graham and G. Lynn Stephens, University of Alabama in Birmingham

"Analyzing Cooperative Computation"  
G. E. Hinton and T. J. Sejnowski, Carnegie-Mellon University and Johns Hopkins University

Paper Session #8

"Long Term Recency Effects in Recalling Previous Answers"  
M. W. Lansdale, ITT Europe

"The Role of Grid Schemata in Memory for Large-Scale Environments"  
Ian Moar, Nancy J. Hamer and Betsy A. Woods, Bucknell University

"The Roles of Interference and Inference in the Retrieval of Autobiographical Memories"  
Brian J. Reiser and John B. Black, Yale University

Paper Session #9

"Thematic Relations Between Episodes"  
Colleen M. Seifert and John B. Black, Yale University

"Reading a Program is Like Reading a Story (Well, Almost)"  
Elliot Soloway, Kate Ehrlich, Eric Gold, Yale University

"Cerebral Lateralization, Preferred Cognitive Mode and Reading Achievement in American Indian Bilingual Children"  
Fairlee E. Winfield, Northern Arizona University



# **Paper Session #1**



## FREQUENCY AND CONTEXT EFFECTS

MARGERY LUCAS  
DEPARTMENT OF PSYCHOLOGY  
UNIVERSITY OF ROCHESTER

Introduction

Lexical access is a multi-component process that involves both the access of a word's orthographic and phonological codes as well as the activation of meaning. The study of this process as it occurs during sentence comprehension is valuable for two reasons. First, we need to know how this important part of sentence processing functions in interaction with other levels of analysis. More generally, its study can provide evidence about how top-down or knowledge-level processing influences processes that occur at lower or stimulus-driven levels. In particular, in this paper I report on a series of experiments concerned primarily with how semantic information as well as information about frequency of use of word meanings affects that part of the lexical process that involves the activation of meaning. For ease of presentation I use the term "lexical access" throughout this paper to refer only to that part of the process.

We know from previous research on lexical access that it is both an autonomous and automatic process. Let us consider the autonomous aspect first.

There are many levels of analysis in sentence processing - the phonological/orthographic, lexical, syntactic, semantic, and pragmatic levels are the ones commonly recognized. These levels must be discrete to be considered as separate modes of analysis but must also interact if a coherent representation of a sentence is to be achieved. How does this exchange of information between levels occur? There are two predominant views on the matter. According to the autonomous model (Garrett, 1978; Forster, 1979; and Swinney, 1979), processing within a level (or "module") is carried on in isolation from processing at other levels. Interaction is limited to the outputs of processing. In the alternative model (Marslen-Wilson and Tyler, 1980) information from higher levels of analysis can facilitate the processing that occurs at lower levels before that processing has been completed.

Evidence for the autonomy of lexical access comes from the study of lexical ambiguity resolution. Many words are associated with more than one meaning. For example, "draft" can mean either "conscription" or "breeze". In a sentence where biasing context precedes the ambiguous word both semantic and syntactic information is available for aiding disambiguation. The crucial question for the autonomy issue is whether this higher-level information acts to restrict access to only the appropriate sense (prior decision hypothesis) or whether it can only have an effect on the process after both meanings are accessed (post decision hypothesis). The former hypothesis is consistent with the interactive model, the latter with the autonomous model.

The work of Swinney (1979) and Tanenhaus, Leiman, and Seidenberg (1979) using cross-modal priming has shown that the post-decision hypothesis is correct. Cross-modal priming is a technique based on the priming paradigm developed by Meyer and Schvaneveldt (1971). They found that the time to decide if a visually presented string of letters is a word (the lexical decision task) is facilitated if a semantically or associatively related word is presented just prior to the target. This result supports a model of semantic memory wherein word meanings are contained in a network of nodes that pass activation along connecting arcs to other, semantically related nodes (Collins & Loftus, 1975). In cross-modal priming the word priming the visual target is presented auditorily. In this way, the priming effect is maintained and lexical access can be studied "on-line", that is, during the actual processing of the sentence. In Swinney (1979), subjects listened to sentences containing ambiguous words in which the sentences were biased towards one or the other meanings of that word. Simultaneous with the offset of the auditorily presented ambiguous word, subjects saw a visual target that was related to one or the other meaning of the word, or a control, as in the following example:

Auditory Stimulus:	The man was not surprised when he found spiders, roaches, and other bugs in the corner of his room.
	↑
Visual target:	ant spy sew

Subjects had to decide if the visual target was a word. Swinney found that targets related to either meaning of the ambiguous word were responded to more quickly than the control target, indicating that both meanings had been activated. Since this occurred even though his sentences were strongly biasing, he concluded that the post-decision hypothesis was correct.

Swinney used noun-noun ambiguities in his study, demonstrating that semantic and pragmatic information could not influence the lexical access process until its outputs were available (4 syllables later only the appropriate meaning was still activated). Tanenhaus, Leiman, and Seidenberg (1979), showed that this was true for syntactic information as well by embedding noun-verb ambiguities (e.g., "rose") in syntactically disambiguating contexts. They used a naming latency task to measure activation. In naming latency, the time to initiate the pronunciation of a visually presented word is measured. They found that visual targets related to either meaning of an ambiguous word were primed even in syntactically disambiguating contexts, again confirming the post decision hypothesis.

Another aspect of the model of lexical access is that it is an automatic process. A number of psychologists have made a distinction between automatic and attentional processes (Kahneman, 1973 and Schiffrin and Schneider, 1977), where attentional processes are controlled, use a pool of limited capacity and are characterized by inhibition of responses that compete with the task at hand. Automatic processes are not under conscious control, do not use capacity, and exhibit facilitation of certain responses but no inhibition. Lexical access appears to be an automatic process since the access of meanings is not under control. Both meanings are always accessed regardless of biasing context. Also, while there was facilitation of responses related to the word in the studies cited above, there was no inhibition of unrelated responses.

Further Determination of the Model

Our knowledge of lexical access is by no means complete. The aspects of the process I undertook to investigate were frequency effects and time course. Previous studies had looked only at the 0 msec. delay and at one delay after word offset. Given how quickly access and decision processes were found to take place, I thought it was necessary to probe at a number of delays to get a more complete picture of the process. In particular, I was interested in the effect of context at different delays. Was it true that context provided no input to the access process until access had been completed? Or was it that its effects were not discernable at the 0 delay? Perhaps probing at the beginning of the word, just at the start of lexical access, might be more revealing of the influence of context than probing at the end of the word when the process was already well under way.

Another issue that concerned me was frequency effects. The meanings of an ambiguous word differ in the frequency with which they occur in language use. For example, the "harbor" interpretation of "port" is more often encountered than the "wine" meaning. Hogaboam & Perfetti (1975) suggested that information about frequency is contained in the lexical network in such a way that dominant meanings are always accessed first and compared with context for appropriateness. Subordinate meanings are only accessed if the dominant meaning is not intended. Hogaboam and Perfetti presented evidence for this but since their dependent measure (reaction times to make judgements of ambiguity) has a strong memory component, it is not clear what the implications are for ongoing sentence processing.

Onifer and Swinney (1981) looked at frequency effects "on-line" using the same paradigm as Swinney (1979) and a 0 msec. delay. They chose ambiguous words that had a clearly dominant (more than 75% frequency of occurrence) and a clearly subordinate (less than 25% frequency) meaning and found no differential effects. Both meanings were accessed even when the sentence was biased toward the dominant meaning. But, although frequency effects weren't found at the offset of the word it is possible that a more thorough investigation of the time course of activation would reveal where and how frequency information is used. One possibility is that dominant meanings reach a greater degree of activation. This would mean that frequency information is encoded in the network in the form of higher resting levels of activation for dominant meanings. This would be reflected in more priming for dominant-related targets. At word offset, both meanings may have just passed their threshold and therefore seem to be equally activated. To determine if this were true, one needs to probe at a later delay.

Yet another hypothesis is suggested by the use of a different type of ambiguous word than has commonly been used in these studies. So far, only homographs (words that sound and are spelled alike but have different meanings) have been used. Another type of ambiguous word was used in the studies to be reported here - non-homographic homophones. These are words that sound alike but have different spellings as well as meanings (e.g., air-heir). Using this type of word has two advantages. It provides a more reliable measure for assessing frequency of occurrence of meanings than the use of homographs. The latter requires word association data - an indirect measure of frequency. Non-homographic homophones enable one to look up the frequencies in the Kucera and Francis (1967) word frequency norms.

A second advantage comes from the additional information that's available in the form of separate orthographic representations for the two words. It's possible that the separate orthographic paths may permit selective access. If one meaning has already been primed by context, access may be restricted to that path with the orthographic representation that is connected directly with that meaning. This mechanism is not available with homographs. There, both meanings are connected with the same orthographic representation (see Figure 1). Alternatively, this information from the orthography of the word may not be used in this way, so that either it is bypassed or else all orthographic paths from the same phonological representation are always accessed.

Finally, I was also interested in how far the automatic activation of meanings extended. Many ambiguous words actually have more than just two meanings. "Draft," for example, has all of the following meanings:

- A. conscription process
- B. breeze
- C. beer
- D. preliminary sketch
- E. payment order

and those are just the noun interpretations. Are the more infrequent meanings activated as well?

In summary, then, I ran the following three experiments:

1) Frequency effects - the cross-modal priming paradigm was used with non-homographic homophones at a delay of 100 msec. from word offset. 100 msec. was chosen since it was expected that multiple access would still be occurring (Tanenhaus et al. (1979) reported that only the appropriate meaning was still activated at 200 msec. after the end of the word) but that the access process would be advanced enough to pick up frequency differences in level of activation if there were any. Alternatively, if separate orthographic representations were permitting selective access, that would be apparent at this delay, as well.

2) Context effects at initial access - design and materials for this are the same as in the previous experiment except that targets are presented at the beginning of the word. This experiment was undertaken to assess the effects of context as lexical access was just starting.

3) Activation of infrequent meanings - this experiment was similar to the above in design but the materials were different. Due to the difficulty in finding three-way ambiguous non-homographic homophones, homographs were used instead. Sentences were biased only toward the dominant meaning and targets were related to either the subordinate meaning or the third most frequent meaning associated with the homograph. Targets were presented at the zero delay to provide the most direct comparison with previous studies that had found multiple access.

### Results

The results are presented in Table I (see next page) as reaction times in msec. Amount of priming is determined by taking the difference between the reaction time to the target and its control.

At 100 msec. multiple access was found for the non-homographic homophones. This indicates that the availability of separate orthographic pathways does not permit selective access and confirms the Onifer and Swinney finding that frequency differences do not lead to selective access. However, at 100 msec. there is a frequency effect in that the degree of priming for dominant meanings is greater in both the appropriate and inappropriate contexts. Apparently, dominant meanings reach a higher level of activation during access than subordinate meanings. This suggests that frequency information is coded in the lexical network in terms of higher resting levels of activation for frequently used meanings.

The results of the second experiment (Targets presented at the beginning of the word) reveal that context is having an effect before the access process has really begun. There is priming for the targets related to the appropriately biased meaning but not for those related to the inappropriately biased meaning. There is no effect of frequency. This demonstrates that context alone has a priming effect. And given that the first study shows that multiple access nonetheless occurs later, one can draw some interesting conclusions about the force of bottom-up processing in lexical access.

Context is able to prime the appropriate meaning of a word before enough phonological information has been obtained to recognize the word. Once the word is heard, however, this effect of top-down processing is overridden by the automatic process of lexical access. Bottom-up analysis apparently has priority here even when there is sufficient information available in the form of strong semantic constraints on probable senses of the ambiguous word.

The finding that context is priming is also interesting in light of the Seidenberg, Tanenhaus, Leiman, and Bienkowski (1982) study that indicated that there are two types of contexts: priming and non-priming. In their experiments, non-priming contexts produced multiple access and priming contexts produced selective access. Priming



contexts contained a semantically or associatively related lexical item prior to the ambiguous word. The results of the present experiments, however, indicate that contexts can be priming yet still produce selective access, suggesting that there was another reason for selective access in the Seidenberg et al. study. Another possibility for this difference in findings concerns the nature of their priming contexts. The contexts in my studies either did not contain words that were strongly semantically or associatively related to the ambiguous word or contained them at a distance from the ambiguous word that was sufficient to make interaction unlikely. Perhaps related lexical items produce activation that is strong enough to overwhelm the automatic access process. In order to assess this hypothesis, one would need to use materials like that of Seidenberg et al. in a condition where targets are presented at word onset to allow a direct comparison with the results of the present study.

The results of the last experiment are problematical. In this experiment, words related to the second and third most used meanings of homographs were presented at word offset in sentences of dominant bias. Results indicate that neither meaning was primed (see Table 2). However, while the difference between related and control words was not significant, it was in the right direction. And, since all previous studies have shown that the subordinate meaning is activated under the conditions of bias that existed in these contexts, I have concluded that there is possibly a confound in this experiment. I am currently collecting subject ratings of the degree of relatedness between targets and their ambiguous words to determine whether poor stimulus-target relatedness obscured the priming effects.

#### Conclusions

The results of these studies taken in sum elucidate the interaction between top-down and bottom-up processing in lexical access as it occurs in sentence comprehension. Biasing contexts can act through the use of semantic constraints to prime anticipated meanings of words. Once the word has been heard, however, an automatic process of access is set into motion in which all the most likely meanings associated with a word are activated regardless of the initial priming due to context. This automatic access of multiple meanings is unaffected by the availability either of information about frequency of use or by potentially disambiguating orthographic information. Once accessed, though, meanings differ in the degree to which they are activated, with the dominant meaning being more strongly activated than the subordinate meaning.

Lexical access, then, is a process in which bottom-up information has priority, overriding the effects of higher-level information and suspending its use until the automatic process set in motion by the auditory signal is completed. Therefore, while top-down processing is essential for sentence understanding, comprehension is also so strongly tied to the stimulus that lower level analyses must be completed before semantic and pragmatic knowledge can have a decisive effect.

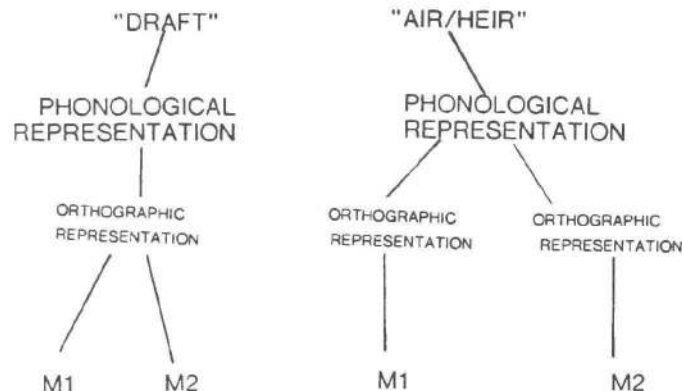


Figure 1

TABLE 1

Delay time : 100 msec

FREQUENCY

	DOM		SUB	
	Related	Control	Related	Control
APP	766.5	831.2	786.3	798.4
	64.7		12.1	
INAPP	789.6	832.3	763.3	790.3
	42.7		27.0	

DELAY TIME : Beginning of Word

FREQUENCY

	DOMINANT		SUBORDINATE	
	Related	Control	Related	Control
APP	782.7	824	811.8	856.0
	41.3		44.2	
INAPP	789.0	775.8	754.0	759.0

TABLE 2

DELAY TIME : 0 mSec

TARGET TYPE

Related M2	Control	Related M3	Control
763	776	787	795

## References

- Collins, A.M., & Loftus, E.F. (1975) A spreading activation theory of semantic processing. Psychological Review, 82, 407-428.
- Forster, K.I. (1979) Levels of processing and the structure of the language processor. In W. E. Cooper & E.C.J. Walker (Eds.), Sentence processing: Psycholinguistic studies. Hillsdale, N.J.: LEA.
- Garrett, M.F. Word and sentence perception. In R. Held, H.W. Leibowitz & H.L. Teuber (Eds.), Handbook of sensory physiology, Vol. VIII: Perception. Berlin: Springer-Verlag, 1978, p. 611-625.
- Kahneman, D. Attention and effort. Englewood Cliffs, N.J.: Prentice-Hall, 1973.
- Kucera, H., & Francis, W. Computational analysis of present-day American English. Providence, R.I.: Brown University Press, 1967.
- Marslen-Wilson, W.D., & Tyler, L.K. (1980) The temporal structure of spoken language understanding. Cognition, 8, 1-71.
- Meyer, D.E., & Schvaneveldt, R.W. (1971) Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. Journal of Experimental Psychology, 90, 227-234.
- Onifer, W., & Swinney, D.A. (1981) Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. Memory and Cognition, 9, 225-236.
- Seidenberg, M.S., Tanenhaus, M.K., Leiman, J.M., & Bienkowski, M.A. (1982) Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. Cognitive Psychology.
- Shiffrin, R.M., & Schneider, W. (1977) Controlled and automatic information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.
- Swinney, D.A. (1979) Lexical access during sentence comprehension. (Re)consideration of context effects. Journal of Verbal Learning and Verbal Behavior, 18, 645-660.
- Tanenhaus, M.K., Leiman, J.M., & Seidenberg, M.S. (1979) Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. Journal of Verbal Learning and Verbal Behavior, 18, 427-441.



# COMPREHENDING AND REMEMBERING NOVEL METAPHORS WHILE READING REAL TEXT

Judith Orasanu

U.S. Army Research Institute

The past five years have witnessed an explosion of research on metaphor. Two critical issues have emerged: Is metaphorical language different in principle from literal language? What is the process of comprehending metaphors?

One extreme position maintains that metaphors are deviant forms of language and therefore require a special comprehension process. According to this view, a word's literal meaning is found to be deviant in the sentence context, is rejected and is subsequently reinterpreted, a two-stage process (C.f., Clark & Lucy, 1975). The other position holds that there is no difference in principle between literal and metaphoric comprehension. Rather, all meaning is constructed based on the text, the context and the reader's pre-existing knowledge (C.f., Rumelhart, 1979). The two-stage model tends to be associated with feature theories of word meaning (e.g., Tversky, 1977; Kintsch, 1974), whereas the single-stage model is associated with schema theory.

Addressing these issues would seem to require that novel metaphors be studied in meaningful context. Unfortunately, few of the studies in the recent literature have done so. The present studies examined the process of comprehending novel metaphors in text and the reader's resulting text concept. One study examined the time course of comprehending metaphors to determine whether readers first access meanings that can be construed as literal or whether they immediately generate contextually-appropriate interpretations. Two other studies examined memory for texts containing metaphors to determine what readers take from the text after they have understood the metaphors.

## Materials

All three studies used metaphors that were found in newspapers or magazine stories or modern novels. Prior to the experiment, groups of adult subjects rated

the metaphoricity, novelty, difficulty, predictability, and imagery levels of over 200 metaphors in context. They also interpreted them. A second group of subjects produced synonyms for the metaphoric words out of context.

From this set we selected a smaller set to use in the experiments. These were items for which there was general agreement on interpretation and a consistent synonym. For example, in the sentence, "The letters made my parents, who are rocks, cry," stoics is an appropriate interpretation of rocks. Out of context, boulders is a high frequency synonym. Metaphors were of two types: nominal (as in the rocks example) and predicative, where a verb is used metaphorically. For example, "My son hurled an obscenity from the bedroom." In this case, shouted is an interpretation, whereas threw is a literal synonym.

### Subjects

Ninety-six adults between 18 and 40 years of age participated in the three studies (ns for studies 1, 2 and 3 were 18, 16 and 64, respectively). They were solicited through an ad in the Village Voice newspaper, and paid for participating.

### Comprehension Study

Procedure. In order to determine whether readers interpret metaphors by first accessing their literal meaning or by directly interpreting the metaphor, a target detection task was used. Targets were metaphors, which occurred in their original contexts (150-250 word excerpts were used). Three types of probes were used: the actual metaphor, its literal synonym and its interpretation. One probe was presented prior to each passage and subjects were instructed to press a key when they detected a word that meant the same as the probe word. Probe type varied within subjects and was counterbalanced across subjects.

Text was presented on a VT52 DecScope video terminal controlled by a PDP-8 computer. A timer was started when the metaphor was displayed and stopped when the reader pressed the key.

If subjects access literal meaning first, their detection latencies should be shorter for synonyms than for interpretations. However, if they directly interpret



the metaphor, this pattern should be reversed.

Results. As shown in Table 1, detection latencies were shorter when probes were literal synonyms than interpretations,  $F(2, 30) = 35.72, p < .001$ . Predictably, identical word matches were fastest. Also, more targets were missed when probes were interpretations than synonyms,  $F(2, 30) = 50.74, p < .001$ .

#### Text Memory Study

Procedure. After finding that readers detect literal aspects of metaphor meaning prior to interpretive aspects, we wanted to determine what information is added to the reader's text concept following the metaphor. A recognition memory procedure was used in which subjects again read text containing a metaphor and judged whether a test word had appeared in the immediately preceding text. Test items were the same words that served as probes in the first study: the metaphor itself, literal synonyms and interpretations. Unrelated words served as controls. Test words were presented at either zero- or five-word lags following the metaphor and within or across clause boundaries.

Results. Recognition errors and decision times are presented in Table 2. Significantly more recognition errors were made to interpretations than to the other three types of test items,  $F(3, 45) = 2.77, p < .05$ . That is, subjects judged that they had seen the interpretation in the text when they had not. Response times were slower for interpretations and synonyms than for metaphor words themselves, but did not differ from each other,  $F(3, 45) = 11.52, p < .001$ . No effects were attributable to lag or sentence structure.

#### Running Memory Study

Procedure. Study three evaluated the contribution of context to the interpretation of the metaphors used in these studies. A running memory procedure used the same metaphor sentences as the previous studies, but out of context. We wanted to determine whether subjects would still confuse test sentences containing the metaphor interpretations with the original metaphors.

Three variants were formed of the metaphor sentences. The metaphor words were

replaced by their interpretations or literal synonyms. The third variant was a control in which one word (not the metaphor) was changed to alter sentence meaning. These variants were substituted for a repetition on half the repetition trials. The number of sentences intervening between initial presentation of the metaphor sentences and test sentences was 1, 4, 9 or 19.

Results. Recognition errors are presented in Figure 1. Again, more false recognitions occurred to interpretations than literal synonyms,  $F(3, 186) = 9.64$ ,  $p < .001$ . There also was an effect of lag: fewer errors occurred with 1 or 4 intervening items than with 9 or 19,  $F(3, 186) = 3.29$ ,  $p < .05$ .

#### CONCLUSIONS

The target detection task showed that readers access the literal meaning of a metaphor before establishing its contextually appropriate meaning. The finding appears to support a two-stage process, i.e., readers access a literal interpretation, find it deviant, and reinterpret the word to fit the context. However, that interpretation may be hasty. We have no evidence that readers find the metaphors deviant. What is clear is that there is a lexical access stage in which literal aspects of meaning are accessed very quickly. Contextually inappropriate aspects of word meaning are temporarily activated, but are not retained. These data argue against parallel access of the literal and interpretive meanings of metaphors, as suggested by Glucksberg, Gildea and Bookin (1982).

They also argue against a strict constructivist approach in which the context so constrains possible word meaning that the interpretation is directly accessed, rendering metaphor interpretation no different from non-metaphor comprehension. Ortony (1979) supported this position, based on his finding that with adequate context, sentential metaphors take no longer to process than their literal equivalents. However, some of his sentential metaphors were actually idioms and may involve different processes than novel single-word metaphors.

The pattern of findings from these three studies is consistent with the notion that an abstract core meaning of words used metaphorically is accessed quickly and

serves as the basis for constructing a contextually-appropriate interpretation, a version of the interaction view of metaphor understanding (C.f., Verbrugge & McCarrell, 1977). Core meaning would presumably relate more strongly to the synonym than to the metaphoric interpretation, yielding the target detection and immediate recognition findings. The present design does not allow us to distinguish between this hypothesis and the two-stage comprehension notion.

#### REFERENCES

- Clark, H.H. & Lucy, P. Understanding what is meant from what is said: A study in conversationally conveyed requests. *JVLVB*, 1975, 14, 56-72.
- Glucksberg, S., Gildea, P. & Bookin, H.B. On understanding nonliteral speech: Can people ignore metaphors? *JVLVB*, 1982, 21, 85-98.
- Kintsch, W. *The Representation of Meaning in Memory*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1974.
- Ortony, A. Beyond literal similarity. *Psychological Review*, 1979, 86, 161-180.
- Rumelhart, D. E. Some problems with the notion of literal meaning. In A. Ortony (ed.), *Metaphor and Thought*. Cambridge: Cambridge University Press, 1979.
- Tversky, A. Features of similarity. *Psychological Review*, 1977, 84, 327-352.
- Verbrugge, R. R. & McCarrell, N. S. Metaphoric comprehension: Studies in reminding and resembling. *Cognitive Psychology*, 1977, 9, 494-533.

Table 1

Target Detection Latencies and  
Misses for Each Probe Type

	Probe Type		
	Metaphor	Synonym	Interpretation
Latencies <sup>a</sup> (in msec)	990	1365	1634
Number of Misses	.42	1.72	3.23

<sup>a</sup>Correct responses only

Table 2

Recognition Errors and Response Latencies  
for Each Item Type

	Item Type			
	Metaphor	Synonym	Interpretation	Unrelated
Errors	3.2	2.5	5.0	.6
Latencies <sup>a</sup> (in msec)	1522	1778	1754	1645

<sup>a</sup>Correct responses only

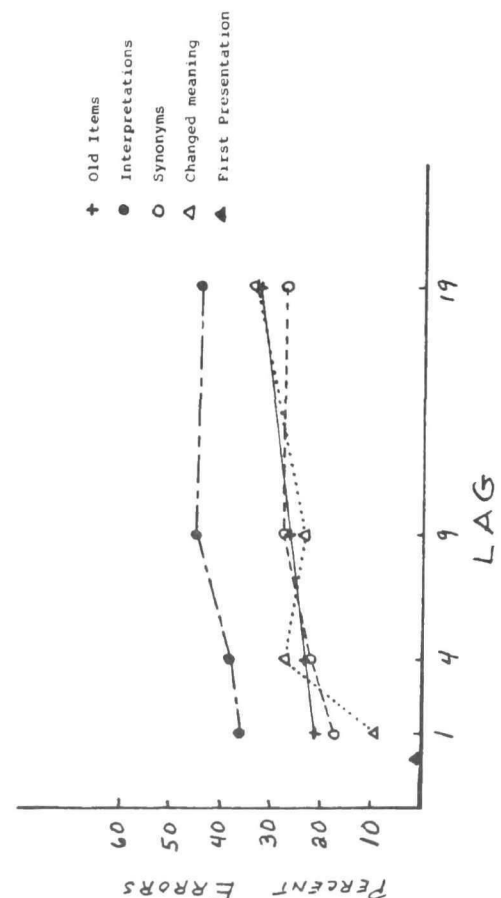


Fig. 1. Recognition errors as a function of item type and lag.



## Verb Semantic Structures and On-Line Language Processing

Ronan G. Reilly  
Educational Research Centre  
St Patrick's College, Drumcondra  
Dublin 9, Ireland

**Abstract:** *This paper describes the use of on-line processing of anaphoric reference to explore the structure of sentence representations. Of central concern is whether or not the representation of verbs is componential. The differential amounts of processing required by subjects to resolve references to objects and simple or complex actions was used to provide insights into the nature of the internal representation of sentences. Two conditions were created; one in which componential effects occurred, and another in which they did not. It was concluded that processing demands dictated the nature of the representation, and that either a wholistic or componential mode of representation could be used on-line with equal facility.*

### Introduction

This paper will examine the issue of semantic decomposition in the context of the on-line processing of sentences. The concept of semantic decomposition has widespread appeal for linguists (Jackendoff, 1976), computer scientists (Wilks, 1977; Schank, 1975), and psychologists (Norman, Rumelhart, and the LNR Research Group, 1975; Gentner, 1981). Gentner (1981) points out that a decompositional approach allows one to account, within one framework, for subjective similarity in meaning, substitutability in paraphrases, and confusability in long-term memory.

However, in spite of its theoretical elegance, the psychological reality of semantic decomposition has not found much empirical support (Kintsch, 1974; Thorndyke, 1975). Gentner (1981) claims that this is due to inadequacies in the type of hypotheses that have been formulated, rather than in the theory itself. The typical hypothesis assumes that a decompositionally complex concept requires more processing to construct and its elements are more difficult to retrieve. Gentner distinguishes between two types of decompositional complexity, connective and non-connective. In a connectively complex representation there is a greater number of semantic relations between the elements representing the nouns in the phrase or sentence. In a non-connectively complex representation there are fewer redundant connections between these elements.

Previous attempts to discover a complexity effect had confounded these two types of representation. To test the distinction, Gentner predicted that cued recall of nouns from a connectively complex representation should be better than from either a non-connectively complex or a simple representation. These predictions were borne out by the results of her experiments.

In light of Gentner's re-definition of decompositional complexity, the purpose of this study was to find evidence for semantic decomposition during the on-line processing of sentences. In keeping with the findings of Tyler and Marslen-Wilson (1977) and Marslen-Wilson and Tyler (1980), it is assumed that all useful sources of information are utilised on-line during the processing of a sentence or phrase. Neither syntactic nor semantic processing are considered autonomous. Both processes interact freely as they continuously update the meaning representation of the linguistic input.

### Anaphora

The processing of anaphoric references was used as a tool to explore the structure of the internal representation of sentences. Linguistically, the anaphora used can be defined as 'surface', rather than 'deep' (Hankamer & Sag, 1976; Sag, 1979). The specific pro-form employed was the word "it". This can be made refer to a particular object, as in the sentence-pair "John read Mary's essay. It annoyed him." Then, simply by changing the final pronoun, thus: "John read Mary's essay. It annoyed her", the referent of the "it" is no longer Mary's essay, but the action of John reading Mary's essay.

The point of resolution of the "it" anaphoric reference in this type of sentence-pair is the occurrence of the final pronoun. Therefore, a measure of the length of time it takes to process the final pronoun should provide a measure of the ease with which the "it" reference is resolved. This in turn should depend on the nature of the representation being accessed in the resolution. If the representation is wholistic and non-verb-central, such as ACT (Anderson, 1976), then there should be no difference in the ease with which object and action references are resolved. If the representation is decompositional, such as MARGIE (Schank, 1975) or the LNR model (Norman, Rumelhart, *et al.*, 1975), it should take longer to resolve an action reference than an object reference. This is because of the diffuse way in which verbs are represented in such a system. Furthermore, if the action referred to is a connectively complex one, it should take longer still to resolve a reference to it. Also, if Gentner is correct, references to objects in a

connectively complex representation should be resolved more rapidly than if the representation were simple.

A third possibility is that the representation of verbs is wholistic, but that they are represented differently from nouns (Kintsch, 1974; Huttenlocher & Lui, 1979). In this case there should be a difference in the ease with which object and action references are resolved, but there should be no effect due to verb complexity. The following experiment was designed to test the above hypotheses.

### Method

**Subjects:** Thirty-one students of St Patrick's College of Education served as voluntary subjects for this experiment.

**Materials and Design:** Eight pairs of core sentences were constructed. They were similar in form to the sample sentence-pair given earlier. The names, actions, objects, and outcomes were varied. However, all sentence-pairs were constructed so that either the pronoun "him" or "her" would be meaningful as the last word of the sentence-pair. From this core of eight, six groups of eight sentence-pairs were generated. They consisted of the following:

(1) Eight sentence-pairs in which the verb in the first sentence was conceptually simple, and in which the "it" in the second sentence referred to the object of the verb. The choice of verb was determined by the scheme proposed in Gentner (1981). Where possible, suitable simple/connectively complex verb-pairs which she used were also employed in this study.

(2) Same as (1), except that the verb was replaced with its connectively complex counterpart.

(3) Same as (1), except that the reference was to the action in the first sentence, rather than the object.

(4) Same as (2) except that the reference was to the action in the first sentence.

(5) Eight sentence-pairs containing a simple main verb, for which the subject was asked to provide either "him" or "her" as the appropriate continuation.

(6) Same as (5) except that the main verb was connectively complex.

In each of the above descriptions "same" means that the

sentences were of the same syntactic form and that they contained the same verb. The names and gender of the individuals were varied, as were the objects of the verb. However, the objects were kept as similar as possible. For instance, "essay" in the sample sentence was replaced by "poem" or "article".

### Procedure

Each subject was required to read all of the 48 sentence-pairs from a computer controlled display. The sentence were presented in two parts. The first part consisted of all of the pair except the final pronoun. In the case of the sample sentence-pair, the subject first saw "John read Mary's essay. It annoyed". After reading this, the subject pressed a key. This cleared the screen, started the timer, and displayed one of these three forms of word combination:

(1) him	(2) hut	(3) her
had	her	him

A schematic representation of the presentation paradigm is given in Figure 1. In the case of word-pair (1) and (2), only one word of each pair could meaningfully complete the sentence. Depending on the gender of the actor in the first sentence the pronoun could cause the "it" in the second sentence to refer either to the object or to the action in the first sentence. In (1) and (2) the alternative word always began with "h" and was never a meaningful continuation of the final sentence.

The purpose of condition (3) was to discover what was the preferred referent in each of the eight simple and eight complex sentence-pairs. It was assumed that by the time the subjects encountered the final pronoun they were already predisposed to a particular referent. Word-pair (3) was designed to provide an approximate measure of this predisposition. This would permit the separation of the subjects' responses into two categories. One containing those responses where the subject had correctly anticipated the referent of "it", and one containing those responses where the referent had been falsely anticipated. Of course, this categorization could only be approximate, since it assumed that subjects' responses to (3) were consistent with their previous anticipations. The usefulness of this categorization will become more obvious in the discussion.

Having decided on the appropriate continuation the subject responded by pressing a key corresponding to either the top or bottom word position. This stopped the timer and caused the next sentence-pair to be

displayed. The process continued until all 48 sentence-pairs had been read. The set of response times (in centiseconds) thus produced were assumed to measure the amount of processing required to resolve the anaphoric references. This was expected to be a fairly accurate measure of the amount of processing required to access the internal representation.

T1	L1	
	2	John analysed Mary's essay. It annoyed
	3	
T2	L1	her
	2	
	3	hut

(T = Time; L = Line.)

**Fig. 1.** A schematic representation of the presentation technique used in the experiment. Only the text of the sentence was visible to the subject.

### Analysis and Results

Each subject contributed  $2 \times 2 \times 8$  response times of interest. These corresponded to the two levels of verb complexity (Complexity) by the two levels reference target (Referent) by the eight sentences. These data were divided into two sets. One set containing those observations where the subject might have correctly anticipated the referent, and one containing those observations where the subject might have falsely anticipated the referent. Membership of these categories was determined by the responses given for sentence types (5) and (6). Each subject's data matrix was then averaged over the sentence dimension. Observations outside the range  $MEAN \pm 2.5 \times SD$ , where the MEAN and SD were calculated over the  $8 \times 31$  observations within each Complexity  $\times$  Referent treatment cell, were excluded from the averaging. This yielded two sets of four observations per subject.

A  $2 \times 2$  repeated measures analysis of variance was performed on each set of observations. Only the Referent factor proved significant in the analysis of the correct anticipation data;  $F(1,30) = 17.81$ ,  $p < .001$ . However, both Referent and Complexity were significant in the analysis of the false anticipation data;  $F(1,30) = 12.69$ ,  $p = .001$  and  $F(1,30) = 5.67$ ,  $p = .024$ , respectively. There were no significant interactions in either analyses. It was considered inappropriate to treat sentences as a random factor because of the way in which they were constructed. Therefore, no quasi-Fs were computed.

The cell means for each analysis are given in Table 1.

Anticipation	Verb Complexity	Referent	
		object	action
Correct	Simple	92.68	105.94
	Complex	92.85	108.34
Incorrect	Simple	100.00	110.47
	Complex	103.51	121.02

**Table 1.** Mean resolution latencies in centiseconds for the 31 subjects.

### Discussion

The false anticipation data from the experiment are assumed to be a more sensitive measure of the processing involved in accessing the internal representation. This is because the subject must resolve the reference from scratch having anticipated the alternative one. The significant Referent effect in these data is an indication that verbs and objects are represented differently. However, almost the same difference between object and action references can also be found in the correct anticipation data (differences of 14 and 15 csecs. for correct and false anticipation, respectively). This means that the Referent effect is not a function of referent anticipation, even though anticipation did have a significant influence on response times, and in the obvious direction (correct anticipation: 101 csecs., false anticipation: 109 csecs.;  $t = 3.12$ ,  $df = 30$ ,  $p = 0.004$ ). The possibility that readers always check for an object reference first, and then process other types of reference is ruled out by the fact that the Referent effect is equally strong under both anticipation conditions. If this heuristic was being used there would either be no Referent effect in the incorrect anticipation data, or a considerably diminished one.

The significant Complexity effect in the false anticipation data is evidence for the componential representation of verbs. However, the effect was not exactly as predicted from Gentner's formulation. It will be recalled that the connectively complex verbs used in this study should have, according to Gentner, increased the accessibility of the elements representing nouns. In other words, object references should have been more rapidly resolved when the representation was complex. The results do not bear this out. Resolutions of object *and* action references were impeded by the complexity of the representation, although it is obvious from Table 1 that the bulk of the Complexity effect is due to the difference between the two kinds of action reference. This



result throws some doubt on the notion of connective complexity.

The disappearance of the Complexity effect in the correct anticipation data and the robustness of the Referent effect in both sets of data would seem to indicate that the resolution of references in the experimental sentences occurs in two stages. In the first stage, prior to the display of the final pronoun, some form of anticipatory processing of the subsequent reference takes place. This processing primarily benefits action references. In the case of a potential reference to a complex action it probably involves the creation of a higher order node in the network to which a link can be subsequently established. Therefore, the representation of a verb can be either componential or wholistic, depending on processing demands. Resolution takes place only after the final pronoun has been encountered, and is differentially difficult. Action reference links are more difficult to establish than object reference links. Hence, the strong Referent effect in both sets of data. This may be due to the differing directionality of links between verb and noun elements in the representation.

In summary, the findings of this experiment support a componential system of representation. The results suggest that the reason why a verb complexity effect has been so elusive is that the task demands of a paradigm dictate the mode of representation used by subjects.

#### References

- Anderson, J.R. *Language, Memory, and Thought*. Hillsdale, New Jersey: Erlbaum, 1976.
- Gentner, D. Verb semantic structures in memory for sentences: Evidence for componential representation. *Cognitive Psychology*, 13, 56-83, 1981.
- Hankamer, J., & Sag, I. Deep and surface anaphora. *Linguistic Inquiry*, 7, 391-426, 1976.
- Huttenlocher, J., & Lui, F. The semantic organization of some simple nouns and verbs. *Journal of Verbal Learning and Verbal Behavior*, 18, 141-162, 1979.
- Jackendoff, R. Toward an explanatory semantic representation. *Linguistic Inquiry*, 7, 89-150, 1976.
- Kintsch, W. *The representation of meaning in memory*. Hillsdale, New Jersey: Erlbaum, 1974.
- Marslen-Wilson, D.W., & Tyler, L.K. The temporal structure of spoken language understanding. *Cognition*, 8, 1-71, 1980.
- Norman, D.A., Rumelhart, D.E., & the LNR Research Group. *Explorations in cognition*. San Francisco: Freeman, 1975.
- Sag, I. The nonunity of anaphora. *Linguistic Inquiry*, 10, 152-164, 1979.
- Schank, R. *Conceptual information processing*. Amsterdam, The Netherlands: North-Holland, 1975.
- Thorndyke, P. Conceptual complexity and imagery in comprehension and memory. *Journal of Verbal Learning and Verbal Behavior*, 14, 359-369, 1975.
- Tyler, L.K., & Marslen-Wilson, W.D. The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior*, 16, 683-692, 1977.
- Wilks, Y. Methodological questions about artificial intelligence: Approaches to understanding natural language. *Journal of Pragmatics*, 1, 69-84, 1977.



## Lexical Access and Serial Order in Sentence Production

Joseph Paul Stemberger  
Carnegie-Mellon University

Serial ordering in the perception and production of language is a phenomenon that is no longer taken for granted in cognitive science. In the serial models that were in vogue until recently, it is a simple enough phenomenon, and a relatively easy thing to encode. In the parallel interactive models that are currently receiving much attention, however, it turns out to be a very difficult thing to encode (McClelland and Rumelhart 1981, Rumelhart and Norman 1982, Dell and Reich 1980). Problems arise in most models despite the relative simplicity of the tasks that they model; in perception, serial ordering is available in the input, while in the production of words in typing and speech, the serial order is available as a schema made available by each accessed word. The ordering of words in sentences is a far more difficult serial-ordering task, because sentences are not stored as units; the speaker cannot take a semantic/pragmatic representation of the sentence and use it to access a schema for that particular sentence. Rather, s/he must use more general strategies for ordering the words. In this paper, I will present one possible solution to the problem of how words are ordered in the production of sentences. There are several distinct problems that must be solved: accessing exactly the right number of words for the sentence, getting those words in the right order, and what to do when a word appears more than once in the sentence. These problems will be discussed in turn.

First, exactly the right number of words must be accessed. This is not guaranteed by the semantic representation. Consider a sentence such as It has a very nice taste/flavor. In this case, there are two words which are virtually synonymous in the context. On the basis of meaning, it is likely that both words could be accessed, and the speaker would say both, one after the other. How can this be prevented? We must hypothesize that the speaker creates a specific number of positions, or SLOTS, for a sentence. We can view words as being associated with these slots, with the prohibition that only one word will normally be associated with a single slot. All the words that are activated by the semantics/pragmatics will be in competition with each other and inhibit each other. The slots provide a source of activation to a word that enables it to overcome the inhibition from other words. There will thus be as many words accessed as there are slots providing this source of activation.

These slots can be used to determine the serial ordering of the accessed words. We can assume that the slots are controlled in a hierarchical fashion that corresponds to surface syntactic phrase structures. For example, there would be units such as S that control the access and ordering of units such as NP and VP. Slots such as N and V control the access and ordering of words. All serial ordering can be derived from such hierarchical structure, in a way similar to the typing

model of Rumelhart and Norman (1982). We must assume that there is a threshold for the execution of a word, and that words reach that threshold one by one; execution in this order constitutes serial ordering. In the Rumelhart-Norman model, words provide a template that determines the activation levels of the letters to be typed. Specifically, the letter at the beginning of the word keeps all later letters inhibited, the second letter keeps all later letters inhibited, and so on. The result is a pattern of activation where the initial letter is highest in activation, and the activation levels decline the later in the word the letter is located. In production, the activation levels of all the letters gradually rise, and each is typed as it reaches a threshold. Thus, the letters are typed in the proper order. This scheme will be adapted for syntax with two changes: control is hierarchical, and the system uses only activation for serial ordering, not inhibition. Some units are differentially activated, and hence reach the execution threshold earlier than others. In execution, a large burst of activation is given to the highest node, S. That node passes activation on to its daughter nodes, but passes the activation at a faster rate to one daughter than the other; usually NP gets the activation at the highest rate in English. NP passes the jolt of activation on to its daughters, with DET getting activation at a faster rate than ADJ, and ADJ at a faster rate than N. These nodes will also pass the activation on to the associated lexical items. Thus, the lexical item that reaches execution threshold first will be the determiner, then the adjective, then the noun, then the verb, and so on. In this way, the lexical items of the sentence will reach the threshold for activation one at a time, in the right order. Feedback from nodes and words to their associated higher nodes is obviously important. This feedback will serve to increase the activation levels of the associated higher nodes, and that extra activation will be passed on to the later daughter nodes. Since a system of this sort must be set up so that the amount of activation passed by a single connection decreases with an increasing number of connections coming into a node (or, equivalently, that the threshold of such a node is increased), it will be the case that a node with many daughters will get less feedback in a unit time than a node with few daughters. As a result, its activation level will rise more slowly and pass on less activation to both higher nodes and later lower nodes. The activation of later nodes, then, and their associated lexical items, is slowed. Thus, through feedback, the system is sensitive to the length of embedded material; later material will be slowed up if there is a lot of material in the subject NP, for example, and will not tend to be executed early.

This leads us to the second problem. Given that a specific number of slots are activated and that all the right words are accessed, how can we ensure that the words become associated with the right slots, so that they appear in the right order? It is necessary and useful to assume that syntactic slots have conditions on them. Each condition can be viewed as a feature that will give a certain amount of activation to specific types of words. This extra source of activation for words will make it more likely that a particular kind of word will be accessed. For example, an N node gives activation to nouns, making it very likely

that a noun will gain enough activation to be accessed, and making it harder for verbs to gain enough activation to be accessed. Other features on nodes probably include tense, aspect, person/number/gender marking, and case. This addition to the model ensures even more accurate access of words, but does not ensure complete accuracy. For example, if two words share many of the same features, such as two nouns do, it does not guarantee that the right word will be associated with the right slot. Further, it does not guarantee that a noun (e.g. destruction) will not be accessed by mistake for one of the noun positions when the target verb is closely related semantically (e.g. destroy).

To ensure absolutely that a word is associated with the proper syntactic slot, we must posit the existence of a very special type of feature. This feature will code words and slots for some type of role (Bock 1982). This role may be meaningful, such as the semantic or syntactic role of the information in question. It is possible, however, that it may just be an arbitrary feature that is assigned to a given chunk of the semantic/pragmatic structure on a nonce basis. In any event, role codes which units go together, including which slot goes with which word. Every unit, whether a word or a slot, must have several copies, differentiated by role. Role will function like any other feature; if the noun slot is role-1, for example, it will give more activation to role-1 nouns than to any other nouns, generally ensuring that the role-1 noun will be accessed in that position. The extra activation to lexical items that comes from this role feature will thus ensure that the words that are chosen will not only be the right number and of the right syntactic categories and inflected in the proper way, but will also be of the right role. Figure 1 illustrates how this scheme might work in the context of a simple sentence like the dog ate the apple; arrows represent connections that pass activation, while filled circles represent inhibitory connections. An associated chunk of material on the semantic level will give some activation to a lot of lexical units. However, the most highly activated units will be those that are connected with the most activated semantic units, here the copies of the words dog and apple; inhibition between pairs of lexical units will remove all other nouns from the competition. Activation from the role units on the semantic level will give the role-1 copy of dog and the role-2 copy of apple the highest levels of activation, and all other copies of these words will be inhibited. In parallel with this activity, syntactic nodes have also been accessed. The role-1 N node will become linked to a role-1 noun, here dog, while the role-2 N node will become linked to a role-2 noun, here apple. These N nodes will be linked to higher syntactic nodes in such a way that the role-1 node and its associated lexical unit will precede the role-2 node and its associated lexical unit. Thus, this system will access the right lexical units, associate them with the right syntactic positions, and assign them the correct serial order.

The last problem to be solved is what happens when a word appears more than once in the sentence. It cannot be the case that a word is simply linked to two slots. In the serial ordering scheme that we are

assuming, such words would wind up being executed only once, and far in advance of where they should be, since the word will simply sum the activation coming from the two slots and reach execution threshold early. It is necessary to assume that there are several copies of the word available, and that different copies will get accessed for the different positions in the sentence. Since we assume that there are several copies available with different roles, this presents no problem. The solution to the previous problem solves this problem as well.

The system as described here does not represent simply an arbitrary solution to the problem of lexical access and serial ordering. Rather, it has been tailored to account for data from errors which occur spontaneously in natural speech. Stemberger (1982) reviews many of the important properties of speech errors. First, errors where the wrong word is accessed but is related to the target word semantically or phonologically generally are constrained by the features on the syntactic slot; i.e., they are of the same syntactic category and are inflected in the same way. Second, words can be accessed correctly but in the wrong syntactic position(s). Such words are generally of the same syntactic category as the words that they replace, but are often not; the semantic activation of such words means that they can occasionally be accessed in a given position without the help of all the features on the slot. Such words tend to be inflected like the target word, however, even if they are of a different syntactic class from the target word. Such accessing errors lead to anticipations, perseverations, and exchanges. Exchanges are viewed as complex errors, where two words are accessed in the wrong slots. It is possible that exchanges occur more than might be expected by chance. If this is the case, then we must assume that the copies of a single word with different roles have greater levels of inhibition between them than nonrelated pairs of words have, making it harder to use a given word twice in the same sentence. If a correct word is accessed in the wrong slot, this will make exchange errors more likely than chance, since it will be harder to access the word in the correct slot as well. Such inhibition will make it most likely that the other correct word will be accessed in the other slot. However, occasionally some other word entirely may be accessed in that slot, generally a word that is related to the target word; such errors are called "bumper cars" by Stemberger (1982). Various other types of errors are also predicted that do occur, such as partial assimilations and sequential blends. As the model predicts, there are also frequent errors, known as "shifts", where two adjacent words flip around and are executed in the wrong order, as one word reaches threshold prematurely. This description of lexical access accounts for a wider variety of the types of speech errors that occur during lexical access and serial ordering than any other model proposed to date.

The problem of accessing words and putting them in the right order is not trivial. I have presented a preliminary solution to the problem, and one that accounts for much of the known data on problems that arise during lexical access. Much work remains to be done on working out the full ramifications of this model, and on implementing it.

# REFERENCES

- Bock, K. Toward a cognitive psychology of syntax: Information processing contributions to sentence formation. Psychological Review, 1982, 89, 1-47.
- Dell, G., and Reich, P. Toward a unified model of slips of the tongue. In V. Fromkin (ed.), Errors in linguistic performance, pp. 273-286. New York: Academic Press, 1980.
- McClelland, J., and Rumelhart, D. An interactive activation model of perception, Part 1. Psychological Review, 1981, 88, 375-407.
- Rumelhart, D., and Norman, D. Simulating a skilled typist: A study of skilled cognitive-motor performance. Cognitive Science, 1982, 6, 1-36.
- Stemberger, J. The lexicon in a model of language production. Unpublished doctoral dissertation, UCSD, 1982.

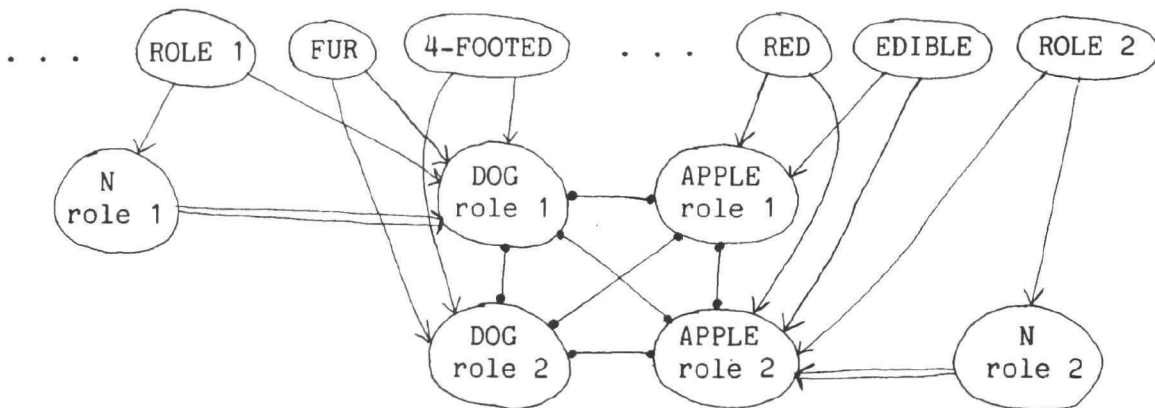


FIGURE 1: Access of nouns in the sentence "the dog ate the apple".



# **Paper Session #2**





# COMPUTATIONAL SELECTION OF PROCESSING LOCATIONS IN VISION

by

Roger Browse

Department of Computing and Information Science  
Department of Psychology  
Queen's University, Kingston, Ontario, Canada

## 1. Introduction

Some of the issues addressed in computational vision research relate to the question of the structure of human intelligence in general, and are thereby subject to a Cognitive Science approach. This is particularly true of the traditional issue of how knowledge of specific objects might be applied towards visual recognition and understanding.

There are two advantages to the development of computer vision systems which maintain compatibility with selected aspects of human vision. First, clues to the underlying operational requirements of the vision system may be obtained from the characteristics of the human system, and second, the resulting computational system may act as an explanatory model for the human operations. The research reported here centers around a computational vision system which interprets line drawings of human-like body forms. This system maintains a number of compatibilities with human vision (see Browse 1981; 1982b).

This paper describes a mechanism for the interpretation-based integration of information available at different levels of resolution. From the integration, there follows a method for the intelligent selection of processing location within the image.

## 2. Multiple Resolution Systems

There is a variety of evidence in favor of approaching vision as a process which operates with information available from several different, but related, levels of resolution. The primate visual system has a distribution and structure of retinal receptors which leads to responses based on different receptive field sizes (see Hubel and Weisel, 1979). While several receptive field sizes may coexist at any specific point on the retina, a near-linear increase in field size (decrease in resolution) is exhibited towards the periphery (Wilson and Bergen, 1979).

Consideration of these aspects of the visual system appears in the formulation of contemporary computationally based theories of vision. Marr and Hildreth (1979) present a mathematical formulation of edge detection in which images smoothed with a variety of Gaussian filters, are convolved with the Laplacian operator. The zero-crossings of these convolutions are representative of the image intensity changes in different spatial frequency channels, depending on the value of the space constant of the Gaussian distribution.

One important theoretical question which confronts the development of such theories which maintain compatibility with human vision is: **What mechanism can provide useful interaction between information from different resolution levels?**

For most theories of computational vision, multiple resolution is a tool in the discovery of context-free image features such as edges. This notion has been developed through the introduction of "image pyramids" (Uhr, 1972; Hanson and Riseman, 1975) which represent an image as several interrelated layers, constructed at different resolutions. The base level is the regular digitized image, and the upper layers are successively smaller images, with pixel values derived by some averaging operation on four (or more) pixels at the level beneath it. A number of processing schemes have been devised to use these structures to aid in the detection of image features. The basic idea behind the use of pyramids is that indications of the existence of a feature may be found in a simple search of a smaller, coarser resolution version of the image. This initial detection can be used to direct the extraction of features from the finer levels (see Tanimoto, 1980).

Marr (1976) takes a similar approach. The process of combining the results from different spatial frequency channels relies on the idea that zero crossings at the same location at different scales are

probably the result of the same underlying physical phenomenon. So whenever the segments obtained from two or more (contiguous) channels agree in both position and orientation, an edge is hypothesized.

Other computer vision research attempts an interpretation-based interaction between levels of resolution, using knowledge of the class of objects which comprise the problem domain. Kelly (1971) devised a system which analyzed coarse level features in the context of what was expected for the outline of a head. Thus the subsequent examination of the fine features was able to ignore the other prominent edges produced by the background. Bajczyk and Rosenthal (1980) have extended the interaction between world knowledge and image hierarchy in an inquiry-driven computer vision system which relies on the natural hierarchical relations of the problem domain. The methods are similar to those proposed by Palmer (1977).

Psychological research supports this idea of the involvement of interpretation in the interaction of resolution levels. Kinchla (1974) and Navon (1977) developed an experimental paradigm in which subjects are presented with the task of processing local and global information at the same time. The result from using this method demonstrate different properties of recognition based on visual information from the two levels. Miller (1981) has presented results which indicate the requirement for a model of perception in which information from different levels of resolution feed into a single decision process which integrates the results.

The approach taken in the body-drawing interpretation system is that separate representations are maintained for information from different resolution levels, and that each one has specialized capabilities for interacting with knowledge of the problem domain.

A second question which relates to the use of multiple resolutions is: **How can locations be selected for the application of high resolution visual processing?**

This question is not only of interest in simulations of human visual processing, but is also a concern to the pragmatics of engineering computer vision systems. The issue of intelligently ordering and restricting the processing of an image becomes more important as larger feature detection masks are being convolved with

images of increasing total number of picture elements.

Saccadic eye movements are the most obvious of the visual selection processes. There are many excellent reviews emphasizing the role of cognition in the selection of fixation locations (see Rayner, 1978). It is generally accepted that expectation or conflict within the ongoing visual interpretation influences the choice of processing area (Loftus and Mackworth, 1978). One important aim of the body-drawing interpretation system was to develop a computational basis for these capabilities.

### 3. System Overview

A computational vision system has been implemented which interprets a class of line drawings of human-like body forms as shown in figure 1. The basic features available to the interpretation processes are representative of three different resolution levels:

- (1) fine resolution: line segments and connections between line segments taken from a 1024x1024 image.
- (2) coarse resolution: axis measurements of blobs detected in a 128x128 image.
- (3) very coarse resolution: measure of the amount of detail that exists in each pixel of an 8x8 representation (shown as 32x32).

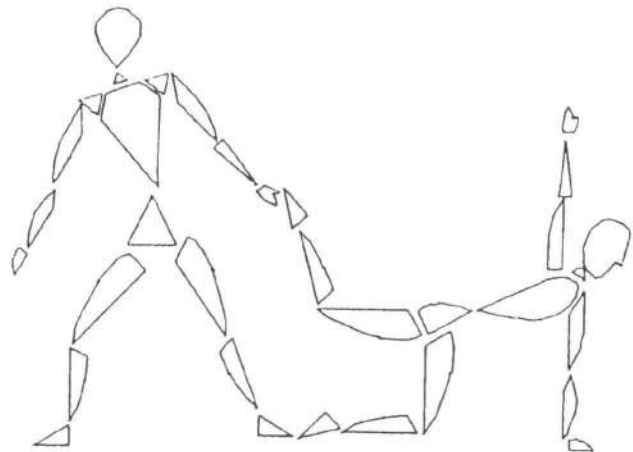


figure 1.

At any given point in the processing, these features are only available in limited diameters of the image, represented as overlapping concentric circles (see figure 2).

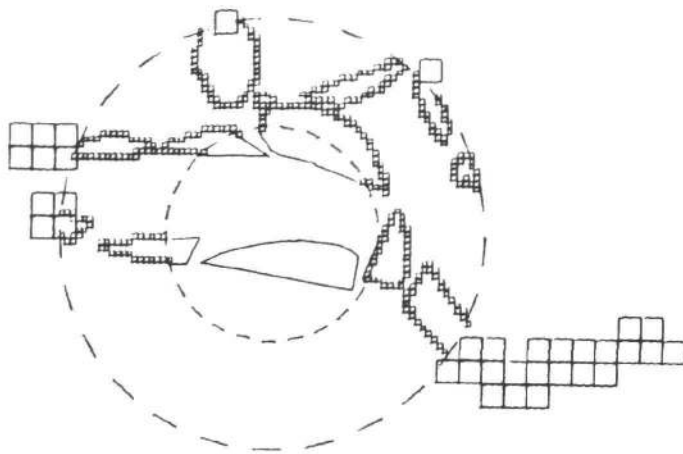


figure 2.

The system uses a schemata-based knowledge representation of the way features may compose to provide supportive evidence for body parts (see Browse 1980, 1982b). This description includes provision for the topologically distinct views of the body parts, and as well encodes the underlying structure of the human body form and its potential for deformations at the joints.

Control of possible interpretations is accomplished by using a cue/model technique adapted from Mackworth's (1977a) MAPSEE system. Through a series of steps involving the application of local consistency methods (Waltz, 1972; Mackworth, 1977b), both at the feature grouping level, and at the model invocation level, interpretation hypotheses are generated regarding the underlying body form being depicted in the image.

#### 4. Interaction Between Resolution Levels

Within limited diameters of available features, complete interpretation of the line drawing is not possible. The results of processing the image during a single fixation (figure 3) is shown in the parse trees of figure 4.

1. Other interpretations exist, providing an ambiguous context. These alternatives have been omitted for the sake of clarity.

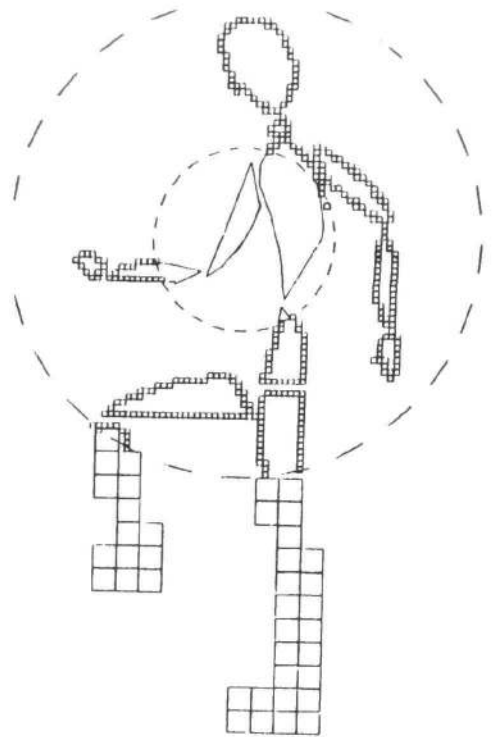
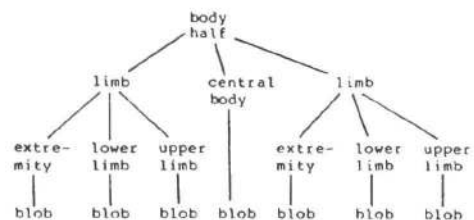


figure 3.



results based on coarse level features



results based on fine level features

figure 4.

These results are available based on both the fine level features and on the coarse level. Note that a more specialized result is produced (in a smaller area) for the fine features.

One obvious objective of the interpretation process is to develop a detailed description for each area of the image. If the system were to rely solely on the processing of fine level features in the accomplishment of this goal, then every location in the image would have to be processed. This exhaustive operation can be avoided through the formation of **correspondences** between interpretations based on the different levels, and by propagating the detailed interpretation outward into the periphery.

Interpretation elements are said to **correspond** if (1) they are related by the specialization hierarchy, and (2) their attributes are similar (particularly location). If a coarse level interpretation object (such as "limb") is known, and one of its components (such as "upper-limb") has a correspondence at the fine level (such as "upper-arm"), then an **inferred correspondence** may be developed for the entire coarse level object: if one of the components has been identified in detail, then the entire structure can be known in terms of its detailed interpretation. These image locations to which detailed interpretation has been propagated may then initiate propagation even further into the periphery.

In order to summarize this mechanism, consider a real-life example: you walk into a room, and your eyes fall on a bookcase in front of you. At that moment, only a few books are actually within the foveal area of detailed vision. The rest of the bookcase is only observed in the lower resolution periphery, and provides an interpretation which may be consistent with several possibilities: records on a shelf, hanging racks of computer tapes, or books on a bookcase. The subjective experience is that of confirmation of "books" over the entire bookcase. The detailed interpretation of "books" at the fovea has propagated a more specific value to the related low resolution interpretation in the periphery.

This notion of correspondence between interpretations is also the mainstay of the system's capability to intelligently select fixation locations. Any coarse

level interpretation object which (1) has components (such as limb does), and which (2) does not have a correspondence at the fine level interpretation, is a prime candidate for a subsequent fixation location. The reason it would be a good candidate is that, if fixated, one of its components will probably be well enough interpreted at the fine level to provide a correspondence adequate to propagate the detailed results to the entire structure. Such locations will provide the greatest possible evidence towards a final interpretation of the image.

The mechanism of consideration of correspondence possibilities provides a means of intelligent selection of fixation location based on the ongoing status of the interpretation process. The selection of fixation locations is also sensitive to properties of the image itself. In the absence of other requirements, locations are selected which are expected to expand the peripheral interpretation area. For the example in figure 3, only three fixations are necessary before an inferred correspondence is established for the entire body form. At that point the system will not have established the exact values of the relative orientations of all body parts, but later fixations may be used to obtain these details as required by the interpretation task.

## 5. References

- Bajcsy, R. and Rosenthal, D.A. 1980, "Visual and Conceptual Focus of Attention," in Structured Computer Vision, S. Tanimoto and A. Klinger (eds.), Academic Press, New York, 133-149.
- Browse, R.A. 1980, "Mediation Between Central and Peripheral Processes: Useful Knowledge Structures," Proc. Third Conf. of the Canadian Society for the Computational Studies of Intelligence, Victoria, Canada, 166-171.
- Browse, R.A. 1981, "Relations Between Schemata-Based Computational Vision and Aspects of Visual Attention," Proc. Third Annual Conference of the Cognitive Science Society, Berkeley, California, 187-190.
- Browse, R.A. 1982a, "Interpretation-Based Interaction Between Levels of Detail," Proc. Fourth Conf. of the Canadian Society for the Computational Studies of Intelligence, Saskatoon, Canada, 27-32.

- Browse, R.A. 1982b, "Knowledge-Based Visual Interpretation Using Declarative Schmata," Ph.D. Thesis, Department of Computer Science, University of British Columbia. Technical Report 82-12.
- Hanson, A.R. and Riseman, E.M. 1975, "The Design of a Semantically Directed Vision Processor," Technical Report No. 75C-1, University of Massachusetts, Amherst, Massachusetts.
- Hochberg, J.E. and Brooks, V. 1978, "Film Cutting and Visual Momentum," in Eye Movements and the Higher Psychological Functions, J.W. Senders, D.F. Fisher, and R.A. Monty (eds.), Erlbaum, Hillsdale, New Jersey, 293-313.
- Hubel, D.H. and Weisel, T.N. 1979, "Brain Mechanisms of Vision," Scientific American 241 (3), 150-162.
- Kelly, M.D. 1971, "Edge Detection in Pictures by Computer Using Planning," in Machine Intelligence 6, B. Meltzer and D. Michie (eds.), American Elsevier, New York, 397-409.
- Kinchla, R. 1974, "Detecting Target Elements in Multielement Arrays: A Confusability Model," Perception and Psychophysics 15, 149-158.
- Loftus, G.R. and Mackworth, N.H. 1978, "Cognitive Determinants of Fixation Location During Picture Viewing," Journal of Experimental Psychology: Human Perception and Performance 4, 565-572.
- Mackworth, A.K. 1977a, "On Reading Sketch Maps," Proceedings of the Fifth International Joint Conference on Artificial Intelligence, Cambridge, Massachusetts, 598-606.
- Mackworth, A.K. 1977b, "Consistency in Networks of Relations," Artificial Intelligence 8 (1), 99-118.
- Marr, D. 1976, "Early Processing of Visual Information," Phil. Trans. Royal Society of London 275B (942), 483-524.
- Marr, D. and Hildreth, E. 1980, "Theory of Edge Detection," Proc. Royal Soc. London B (207), 187-217.
- Miller, J. 1981, "Global Precedence in Attention and Decision," Journal of Experimental Psychology: Human Perception and Performance 7, 1161-1174.
- Navon, D. 1977, "Forest Before Trees: The Precedence of Global Features in Visual Perception," Cognitive Psychology 9, 353-383.
- Palmer, S.E. 1977, "Hierarchical Structure in Perceptual Representation," Cognitive Psychology 9, 441-474.
- Rayner, K. 1978, "Eye Movements in Reading and Information Processing," Psychological Bulletin 85 (3), 618-660.
- Tanimoto, S.L. 1980, "Image Data Structures," in Structured Computer Vision, S. Tanimoto and A. Klinger (eds.), Academic Press, New York, 31-55.
- Uhr, L. 1972, "Recognition Cone Networks that Preprocess, Classify, and Describe," IEEE Transactions on Computers 21, 758-768.
- Wilson, H.R. and Bergen, J.R. 1979, "A Four Mechanism Model for Threshold Spatial Vision," Vision Research, 19, 19-32.





# Topological Invariants and Apparent Motion

Lln Chen

*China University of Science & Technology  
Hofei, Anhui Province  
The People's Republic of China*

and

*Program in Cognitive Science  
University of California at Irvine  
Irvine, CA 92717*

## INTRODUCTION

As many authors (Ullman, 1979; Anstis, 1979; Marr, 1982) have pointed out, at the core of how to understand apparent motion lies the correspondence problem: in the process of perceiving apparent motion, one has to establish at some level a correspondence identifying the parts of two succeeding stimulus frames which represent the same object.

One distinguishing fact of apparent motion is that when one perceives apparent motion, often he also perceives some sort of transformation from one stimulus pattern to another; for example, a square is moving and changing shape simultaneously to become a circle. Kolers & Pomerantz (1971) found that three kinds of transformations, not only translation and rotation of rigid shape but also intriguing plastic deformation, occurred when the dissimilar pairs gave apparent motion. The interesting question is: What kinds of invariants under the transformation of "plastic deformation" does the visual system depend on to determine that two figures, however shape-changed they may be, nevertheless represent the same object?

As Chen (1980, 1981, 1982a,b,c,d) has argued, a primitive and general function of the visual system may be the perception of global topological invariants, such as connectivity, closure and holes, which are defined as invariants under topological transformations (plastic deformations without breaking and fusion). Many paradigms and approaches in visual perception, such as the object superiority effect (Weisstein & Harris, 1974; McClelland, 1978; Williams & Weisstein, 1978; Chen, 1982c), grouping (Olson and Attneave, 1970; Pomerantz, Sager & Stover, 1977; Chen, 1982a), card sorting (Palmer, 1978), effortless texture discrimination (Julesz, 1981), visual sensitivity to distinction made in topology (Pomerantz, 1980; Chen, 1982b) and competing organization with several simultaneous factors (Chen, 1982d) have provided some evidence for detection of topological properties in visual perception. The many facts mentioned above also lead us to consider topological invariants as candidates for the correspondence tokens in apparent motion.

## METHOD

The general method for the following experiments was advanced by Ullman (1979) and is called "the competing motion technique". In this method two stimulus displays are successively presented. The first one contains a single figure in the center, while the second contains two figures located on either sides of, and at the same distance from, the center. The question asked is whether the figure in the first stimulus display is perceived to move to one or the other of the two figures in the second. In Ullman's demonstrations, each set of two stimulus displays were alternated repeatedly, and without providing precise data, he reported the subjects' motion preference. For collecting accurate data, in the following experiments each set of two stimulus displays was presented on just one cycle for each trial, but many trials were used as experimental presentations. This method provides us an experimental measure to compare and characterize the effect of various structural invariants.

A three-field tachistoscope was used for presenting stimuli. Subjects were asked to look at a fixation point at the center of the preexposure field then press a button which resulted in a presentation cycle. Each first stimulus display containing a middle figure of each pair was presented for 100-150 msec. and each second stimulus display, for 1000 msec., with a inter-stimulus interval (ISI) of 20-30 msec.. For each presentation subjects were required to choose in a forced choice procedure one of two responses: "right" (motion from a middle figure to a figure at right) or "left" (motion to a figure at left), guessing if necessary. For each subject the presentation durations were adjusted in order to produce strong effects of apparent motion. Each subject was initially familiarized with the phenomenon of apparent motion under the condition of single alternate exposure. At least three trials of each pair were used as a practice presentation. The order of presentations was randomized and counterbalanced across subjects. Four blocks of 21 trials per block, which contained three presentations of each pair, were used for test presentation.

Four subjects participated in all these experiments involving seven pairs of stimulus displays.

Pair 1 (adapted from Pomerantz et al, 1977) consists of the two stimulus displays shown in Fig.1. The first contains an arrow (stimulus a). The second contains two figures which are made up of exactly the same three line segments as the arrow. The difference between stimulus b and c is just that the one of the shorter line segments is located in two different positions, displaced by a constant distance from the same line segment in the arrow. But the closed nature of a triangle makes it topologically different from the other two figures.

Pair 2 is shown in Fig.2. Each of the three figures is made up of five line segments with two sorts of lengths. Among the three figures are stimulus a and b adapted from Julesz (1980). Although stimulus a has a different number of "terminators" from stimulus c and the same number of terminators as stimulus b, stimulus a possesses the same topological invariant, simple connectivity, as stimulus c and is different in topological invariants from stimulus b, which is disconnected with a closure.

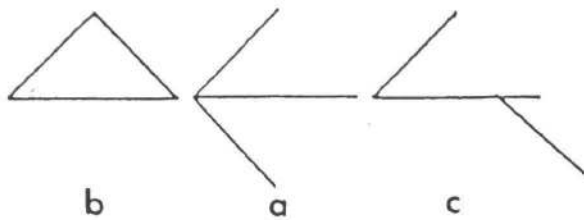


FIGURE 1

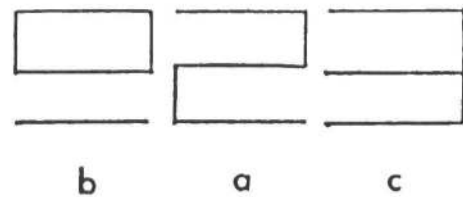


FIGURE 2

In Pair 3, stimulus a is a solid circle, stimulus b is a ring and stimulus c, a solid square. From our intuitive experiences, a solid circle seems to have more "similarity" to a ring than to a square; however, from topology, the difference between a disk and a ring is much deeper than that between a disk and a solid square. For the latter, the difference will dissolve under a topological transformation, a plastic deformation without breaking and fusion. On the other hand, stimulus b is characterized as a connected component with a hole in it, a typical topologically invariant description.

Fig.4 shows Pair 4, which is similar to Pair 3 except that a square with a square hole in it was used instead of the disk, the middle figure in Pair 3.

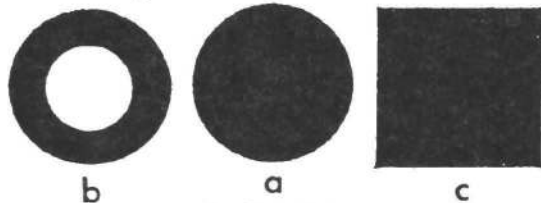


FIGURE 3



FIGURE 4

The three figures contained in Pair 5 have the common feature of possessing holes in a connected component (shown in Fig.5). However, stimulus b has two holes in it, while stimulus a and c, just one. The number of holes is a topological invariant, "the order of connectivity". So, stimulus a is topologically equivalent to stimulus c, but not to stimulus b. The inner diameter of stimulus c is 0.71 ( $\sqrt{2}/2$ ) times that of stimulus b, so the total area of the two inner circles of stimulus b is equal to that of the inner circle of stimulus c.

Pair 6 (shown in Fig.6) involves another kind of global topological invariant, whether a target line is within a closed curve or outside it. If the two stimulus displays were superimposed, both stimulus a (a line segment) and the target line in stimulus c would lie within the big circle of stimulus c, while stimulus b (another line segment) would be outside the circle.

Discrete dots were used to make up Pair 7 (shown in Fig.7). These figures are adapted from Zeeman (1965). Each figure is a set of discrete dots, but globally they look like Pair 3. Even though at the viewing distance in the experiment one can see that these stimuli are discrete dots, one still has a wholistic perception, that is, stimulus a looks like a solid circle; stimulus c, a solid square; stimulus b, a ring. For a clearly discrete set, why do we have continuous and wholistic perception? This fact suggests that the visual system can ignore details within a certain range for attaching importance to global structure (Chen, 1980, 1981).

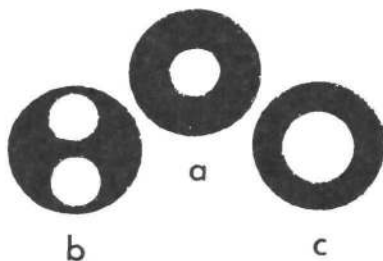


FIGURE 5

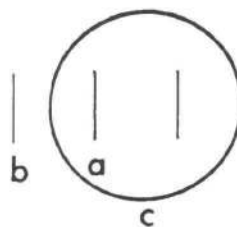


FIGURE 6

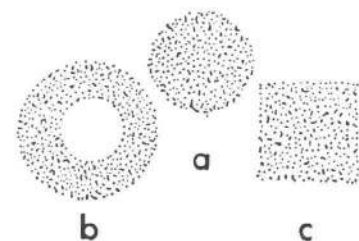


FIGURE 7



## RESULTS AND DISCUSSION

	Subject 1	Subject 2	Subject 3	Subject 4	Average
Pair 1	92%	83%	75%	75%	81%
Pair 2	83%	92%	83%	83%	85%
Pair 3	83%	92%	92%	83%	88%
Pair 4	92%	83%	92%	92%	90%
Pair 5	83%	100%	83%	92%	90%
Pair 6	92%	92%	83%	83%	88%
Pair 7	83%	92%	83%	75%	83%

TABLE 1. The percentages of reporting motion from each middle figure to the figure with the same topological invariants. The difference in average percentages between figures with the same topological invariants and figures with different ones are all significant,  $p < .01$ .

The results with all seven pairs are shown in Table 1. They clearly show that subjects saw strong preference for motion from a middle figure to a figure that has the same topological invariants as the middle figure. The figures used in the first pair were made up of exactly the same line segments. According to the theory (Ullman, 1979) considering line segments as correspondence tokens, the only difference between stimulus b and c is that one component line segment was translated with a constant distance. It seems difficult for this kind of theory to interpret the preference for motion from stimulus a to stimulus c.

In Marr's "primal sketch" (1978) and Julesz' theory of effortless texture discrimination (1981), terminators are also considered as basic primitives. Along this line, somebody might like to argue that the preference for motion between stimulus a and c in Pair 1 was observed not because of closure but just because of terminators. However, the result with Pair 2 seems to rule out this objection. Stimulus a possesses the same number of terminators as stimulus b but not stimulus c. In some cases, terminators may reflect closure and connectivity properties, but not always.

Assuming that the visual response to closure resulted from topological structure in visual perception, it would be predicted that subjects would see motion preferentially between a disk and a solid square rather than between a disk and a ring. The result with Pair 3 supports this prediction.

It is interesting to note that for Pair 3 and 4 only the stimulus in the first display changed; yet the preferred direction of apparent motion strongly reversed from the solid square in Pair 3 to the ring in Pair 4. This pattern strongly suggests topological explanation. In fact, Ramachandran, Anstis and Ginsburg (1982) have already reported that subjects display preference for motion from a solid square to a solid circle rather than an outline square and from a cross to a rotated cross rather than a square made by the same line segments. The facts have been interpreted assuming that "low spatial frequency dominates apparent motion". Along this line of thinking, the preference for motion from a hollow square to a ring would have to be interpreted as high spatial frequency domination of apparent motion. These two observations, therefore, cast some doubt on the value of explanations of apparent motion using the notion of spatial frequency. Under plastic deformation of stimuli, like lines or blobs, spatial frequencies, whether low or high, would be difficult to imagine as invariants for the correspondence tokens.

Because the number of holes in a connected component represents a typical topological invariant and its intuitive meaning is not obvious, Pair 5 was designed to give more evidence for the topological hypothesis and also to further rule out some other factors that are often relevant to the study of brightness sensitivity or spatial frequency analysis. The fact that the total area of the two inner circles in stimulus b is equal to that of the bigger inner circle in stimulus c makes the explanation in terms of differences in brightness or spatial frequency difficult; however, the result is consistent with the topological explanation. One might argue that the motion from stimulus a to stimulus c arises from the "similarity" between them. However, in Pair 3, the motion from the disk to the square rather than to the ring already indicates that the correspondence tokens are not these kinds of similarity factors.

Some long-standing debates about the nature of apparent motion have often centred on the fundamental question of how to understand grouping (Anstis, 1979). Chen (1980, 1981) has considered grouping also as the extraction of global topological invariants. The very nature of visual perception is discrete, and the approach of perceptual organization, say, the Gestalt laws, is often aimed at some obviously discrete stimuli, such as dot arrays. Therefore, general topology cannot be directly used to describe perceptual organization. For a clearly discrete set, why do we have continuous and wholistic perceptions? In this sense, the Gestalt perceptual phenomena look puzzling. The mathematics of Tolerance Spaces (Zeeman, 1962) tells us how to formulate the global properties on a discrete set. Tolerance is an algebraic relation chosen not only to represent the concept of the least noticeable difference but also to represent a minimum measure within the range of which details will be ignored by the perceptual system for attaching importance to global properties.

In a tolerance space, we can build up a mathematical structure similar to topology. Grouping represents a visual function to ignore details within a certain tolerance and to extract global tolerance invariants, such as tolerance connectivity, closure and holes. Taking a tolerance of one centimeter, the two most noticeable global tolerance properties of the second stimulus of Pair 7 are that it has two pieces (two tolerance connected components) and one of them has a hole in it. So, if grouping can really be considered as the extraction of global tolerance invariants, then preference of apparent motion should be observed from the tolerance circle to the tolerance square rather than to the tolerance ring, the global tolerance property of which is different from that of the others. The result with Pair 7 helps us with the suggested theoretical basis for understanding grouping. Correspondence processes with either grouping or normal patterns can be described consistently in terms of topological invariants.

### SUMMARY AND GENERAL DISCUSSION

Using the adapted "competing motion technique", seven pairs of stimulus figures showing topological variation were designed to reveal some evidence for that topological invariants play a role in correspondence processes of apparent motion. These experimental data, which show that subjects reported strong preference for motion from a central figure to a figure with the same topological invariants as its, came from various kinds of stimulus patterns that represent quite different structural forms and that control other explanatory factors, such as brightness, spatial frequency and terminators. Nonetheless all of them are consistent with the topological explanation and strongly suggest a topological structure in visual perception.

The key point is that the units of figure perceptual representation are invariants at different geometrical levels (Chen, 1981). Along this line we can deepen our comprehension of some of the long-standing debates about apparent motion. For example, now it is clear that the question of whether motion perception precedes form perception is not a good question. Form perception includes different levels and at the level of extraction of topological invariants it precedes motion perception; however, at the levels of more detailed properties, say, the difference between a square and a triangle, motion perception may precede form perception. So, we cannot simply claim that "the correspondence tokens are not structured forms" (Ullman, 1979). The right question is which kinds of structured forms should be considered as the correspondence tokens, and which kinds are not. In fact, topological invariants are a kind of important structured form.

The critical act in formulating Ullman's computational theory for apparent motion is of using the rigidity constraint on the way the world behaves (Marr, 1978). Ullman's theory is noted for the discovery of a valid constraint of rigidity, which "enables us to solve the structure-from-motion problem unambiguously" (Marr, 1982). But at the same time a certain limitation of the theory also comes from rigidity. Plastic deformation is a strong and common phenomenon in apparent motion. Marr (1982) pointed out that a new theory may be needed for when the object is not only moving but also changing. The experimental facts reported in the present paper and their topological explanation have suggested a new type of analysis of apparent motion, which has been motivated by the general transformation, plastic deformation.

The suggested topological approach, supported by empirical data, has also raised some interesting issues about Marr's "primal sketch". "The primal sketch" has emphasized "the local geometry of an image". How should we consider the relationship between topological properties and "the primal sketch"? It seems that it is implausible mathematically to compute out topological properties from local geometrical properties, such as oriented edges, lines, blobs. And the fact that perception of topological invariants precedes motion perception has shown the early extraction of topological properties. Many other experiments, for example, the configural superiority effect (Chen, 1981) and competing organization with several simultaneous factors (Chen, 1982d), have also provided some evidence for the extraction of global topological invariants earlier than that of local geometrical properties, such as orientations, positions. So, considering the time dependence of perceived properties, it seems difficult to assume that topological properties are derived from local geometrical properties. Minsky and Papert (1972) proved that for perceptrons, the topological predicate is not finite order; however, lower-order perceptrons can be used for computing geometrical properties. Thus from the perspective of computational theory, the global nature of topological properties makes them essentially different from local geometrical properties. It therefore seems difficult for "the primal sketch" to accommodate topological properties without changing its local nature. These questions are fundamental for understanding vision and deserve further study.

#### Note about figures

All stimulus figures used in these experiments are black on white paper. Line segments in Pair 1, 2 and 6 were drawn by a pen with line width in .50 mm.

Two of three line segments of each stimulus figure in Pair 1 have equal length in 24 mm, the other, 32 mm. The distance between stimulus b and c is 36 mm.

Three of five line segments of each stimulus figure in Pair 2 have equal length in 26 mm, the others, 11 mm. The distance between stimulus b and c is 34 mm.

In Pair 3, the diameter of the disk is the same 32 mm as the outer diameter of the ring and one side length of the square. The inner diameter of the ring is 18 mm. The distance between the ring and the square is 40 mm.

In Pair 4, one outer side of stimulus a is the same 27 mm in length as one side of the square and the outer diameter of the ring. The inner side of stimulus a is 14 mm in length. The inner diameter of the ring is 18 mm. The distance between the square and the ring is 40 mm.

The outer diameters of three stimuli in Pair 5 are all the same 40 mm. The diameter of the inner circle in stimulus a is the same 15 mm as that of one inner circle of stimulus b. The inner diameter of stimulus c is approximately 21 mm. The distances between stimulus a and b and stimulus a and c are all 2 mm.

The lengths of three line segments in Pair 6 are all the same 21 mm. The diameter of the circle is 59 mm. The distance between stimulus a and b is the same 25 mm as that between stimulus a and the target line in stimulus c.

In Pair 7, the diameter of stimulus a is 33 mm, the outer diameter and the inner diameter of stimulus b are, respectively, 43 mm and 19 mm, and one side length of stimulus c is 30 mm. The distances between stimulus a and b and stimulus a and c all 3 mm.

## REFERENCES

- Anstis, S. M. Apparent movement. In R. Held, H.W. Leibowitz, H.-L. Teuber(Eds.), *Handbook of Sensory Physiology vol. 8, Perception*, New York: Springer-Verlag, 1979.
- Chen, L. Tolerance spaces and Gestalt. paper presented at the 4th Biophysical Academic Conference of China, Beijing(1980).
- Chen, L. Perceptual organization and tolerance spaces. paper presented at the 14th Annual Mathematical Psychology Meeting, Santa Barbara(1981).
- Chen, L. What are the units of figure perceptual representation? *Studies in Cognitive Science*, UCI, No. 22, 1982a.
- Chen, L. Topological structure in visual perception. *Science*, 218, 669-700, 1982b.
- Chen, L. Connectedness and the object superiority effect. *Investigative Ophthalmology and Visual Science*, 22, 124, 1982c(Supplement).
- Chen, L. Competing organization with several simultaneous factors. *Bulletin of the Psychonomic Society*, September 1982d, 145-146(abstract).
- Julesz, B. Figure and ground perception in brief presented isodiot-pole texture. In M. Kubovy & J.R. Pomerantz(Eds.), *Perceptual Organization*, Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publisher, 1981.
- Julesz, B. Textons, the elements of texture perception, and their interactions. *Nature*, 1981, 290, 91-97.
- Koler, P.A. & Pomerantz, J.R. Figural change in apparent motion. *Journal of Experimental Psychology*, 1971, 87, 99-108.
- Marr, D. Representing visual information. *Lectures on Mathematics in the Life Science*, vol.10, 101-180. reprinted in Hanson, A.R. & Riesman, E.M.(Eds.) *Computer Vision Systems*. New York: Academic Press, 1979.
- Marr, D. *Vision*. San Francisco: Freeman, 1982.
- McClelland, J.L. Perception and masking of wholes and parts. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, 4, 210-229.
- Minsky, M. & Papert, S. *Perceptrons*, MIT Press, 1969.
- Olson, R.K. & Attneave, F. What variables produce similarity grouping? *American Journal of Psychology*, 1970, 83, 1-23.
- Palmer, S.E. Structural aspects of visual similarity. *Memory and Cognition*, 1978, 6, 91-97.
- Pomerantz, J.R., Sager, L.C. & Stoeber, R.J. Perception of wholes and their component parts: some configural effects. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, 3, 422-435.
- Pomerantz, J.R. Are complex visual features derived from simple ones? in E.I.J. Leeuwenbert & H.F.M. Buffart(Eds.), *Formal Theory of Visual Perception*, Sussex, England: Wiley Ltd., 1980.
- Ramachandran, V.S. Perception of apparent motion. *Studies in the cognitive Sciences*, No. 11, UCI, 1981.
- Ullman, S. *The Interpretation of Visual Motion*. The MIT press, 1979.
- Weisstein, N. & Harris, C.S. Visual detection of line segments: an object-superiority effect. *Science*, 1974, 186, 725-755.
- Williams, A. & Weisstein, N. Line segments are perceived better in coherent context than alone: an object-superiority effect in visual perception. *Memory & Cognition*, 1978, 6, 85-90.
- Zeeman, E.C. The topology of the brain and visual perception. in Fort, K.(Ed.) *Topology of 3-manifolds*. Prentice Hall, Englewood Cliffs, N.J., 1962.
- Zeeman, E.C. & Bunceman, O.P. Tolerance spaces and the brain. In *Mathematics and Computer Science in Biology and Medicine*. (Proceedings of Conference held by the Medical Research Council in Association with the Health Departments, Oxford, 1964).

ACKNOWLEDGEMENTS. I am grateful to William Batchelder for valuable suggestion, helpful discussion and careful review of drafts of this paper. In particular, I would like to thank Paul Smolensky for stimulating comments and reviewing this paper and to thank Bing Chen for helpful suggestion and helping with preparing these experiments. Helpful comments came from Michael Braunschtein, Donald Macleod, Donald Norman, V.S. Ramachandran, David Rumelhart and Jack Yellot. Finally, I would also like to thank Eileen Conway for preparing the manuscript. This research was supported by a grant from Sloan Foundation to Program in Cognitive Science at UCI.



## Imagined spatial transformations in the visual discrimination of left and right parts of the human body.

Lawrence M. Parsons

*Institute for Cognitive Science  
University of California, San Diego*

Much research has demonstrated that to discriminate between pairs of differently oriented but very similar objects, one very often finds it necessary to imagine (or physically produce) the rotation of the two objects into a similar alignment to compare their shapes (Cooper and Shepard, 1982; on physical rotations, Parsons and Hinton, Note 1.) An important property of the ability to imagine the rotation of an object is that often the time required for such a rotation is a linearly increasing function of the extent of rotation.

Most of this research has been done either with abstract shapes that are unfamiliar to subjects or with alphanumeric characters. There has been little investigation of the effect of orientation on the discrimination of more naturalistic stimuli. A few (mostly recent) studies of the perception of natural objects have examined mirror-image discrimination tasks in which subjects apparently imagined a human body or body-part at the orientation of a stimulus (Cooper and Shepard, 1975; Parsons, 1983, Notes 2-5, and 7). This paper reports on a few results of my own investigations in this area (Parsons, 1983, Notes 2-5, and 7).

Participants in my experiments made left-right judgments of variously oriented hands and feet, and of the raised arm of a human body depicted from many perspectives. Figure 1 shows examples of the kind of stimuli used in these experiments.

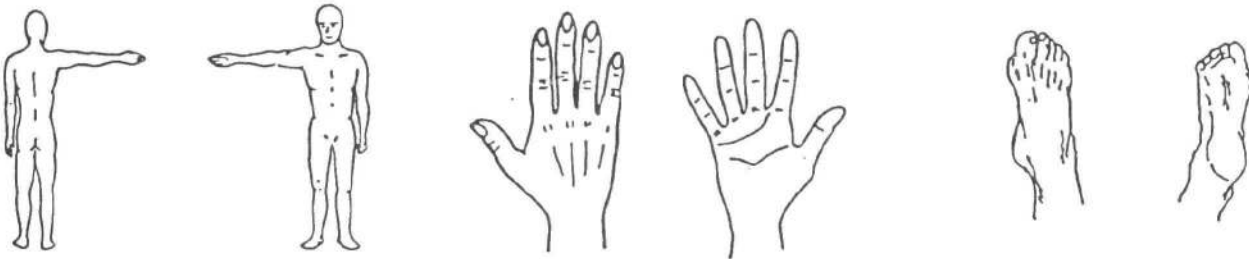


Figure 1.

This task requires button press responses with the left hand for a stimulus portraying a part of the left half of the body and with the right hand for a stimulus portraying a part of the right half of the body. Subjects in various experiments viewed a set of stimuli which depicted either a human body (with a raised arm), a hand, or a foot from many different perspectives.

A major feature of this research is the examination of subjects' abilities and preferences for selecting planes or axes for imagined rotations of objects when the number and variety of possibilities is relatively unconstrained by experimental design. In nearly all previous research, one or the other of two planes or fixed axes of rotation was most efficient to "correct" for the difference in orientation between the standard and comparison objects. Most studies have used orientation differences (ODs) in one or both of two planes (the "picture" and/or "depth" planes); and trials have frequently been blocked by the plane of the OD. By contrast, on any trial in the experiments reported here, any one of many different planes (or fixed axes) of rotation could be uniquely most efficient to bring the standard and comparison object into

Portions of this work were part of a doctoral dissertation submitted to the Department of Psychology, University of California, San Diego. I would like to express my appreciation to Geoff Hinton, Jay McClelland, Don Norman, and Dave Rumelhart for their very generous encouragement, support, and discussion. This research was conducted under National Science Foundation Grant BNS 79-24062 to James L. McClelland.

Send requests for reprints to: Dr. Larry Parsons; Institute for Cognitive Science, C-015; University of California, San Diego; La Jolla, CA. 92093.



congruence. In addition, the orientation difference between standard and comparison objects in these experiments could be considerable. For instance, they could differ simultaneously in the orientation of their "front-back", "top-bottom", and "left-right" aspects. For these kinds of OD, an object can be brought into congruent alignment with another object (i) by being sequentially rotated about a different fixed axis for each dimension by which the two objects differ in orientation; or (ii) by being "simultaneously" rotated about different axes, producing rotations which are effectively about unfixed axes (i.e., about an instantaneously changing axis); or (iii) by being rotated about a fixed axis which simultaneously corrects for all differences in orientation.

Thus, in this paradigm, to "correct" for orientation differences, subjects must select planes (or axes) of rotation from among a relatively large set of possibilities, according to their own criteria of efficiency. Their solutions to these geometrical problems provide new information about basic aspects of the imagination and spatial reasoning ability.

In addition to investigating spatial reasoning abilities in this context, I have been studying in detail how subjects make these seemingly simple and mundane discriminations; further, because subjects (as it turns out) do seem to imagine the rotation of parts or all of their own body in this task, I have been comparing the imagined transformations to corresponding physical actions (Parsons, 1983, and Notes 2-7).

Figures 2-4 show the reaction time-orientation difference (RT-OD) functions for subjects' judgments of whether a particular stimulus body part (a raised arm, a hand or foot) is of the left or right half of the body. Each RT-OD function is an average of results in 2 to 5 different experiments, using different subjects. Each stimulus in Figure 1 was presented in random ordered trials at 30 degree intervals in the "picture" plane, beginning at the position (0 degrees of OD) shown in Figure 1.

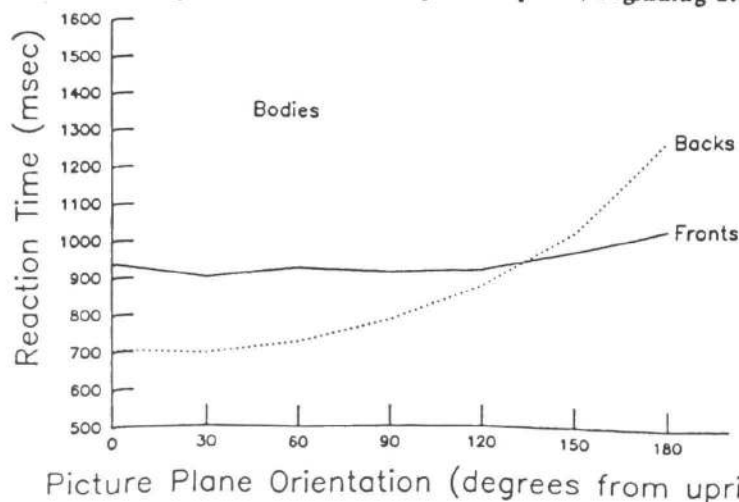


Figure 2. RT-OD functions for fronts and backs of bodies. Based on 4610 correct responses of 31 subjects.

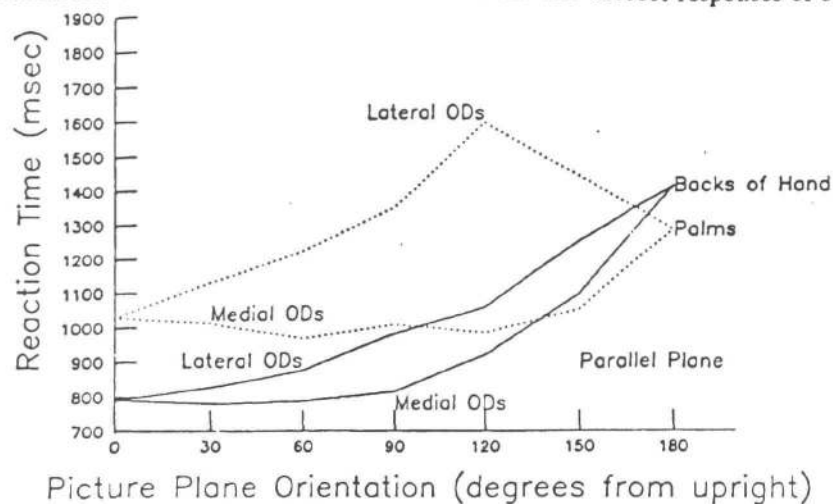


Figure 3. RT-OD functions for palms and backs of hands. Based on 4992 correct responses of 19 subjects.

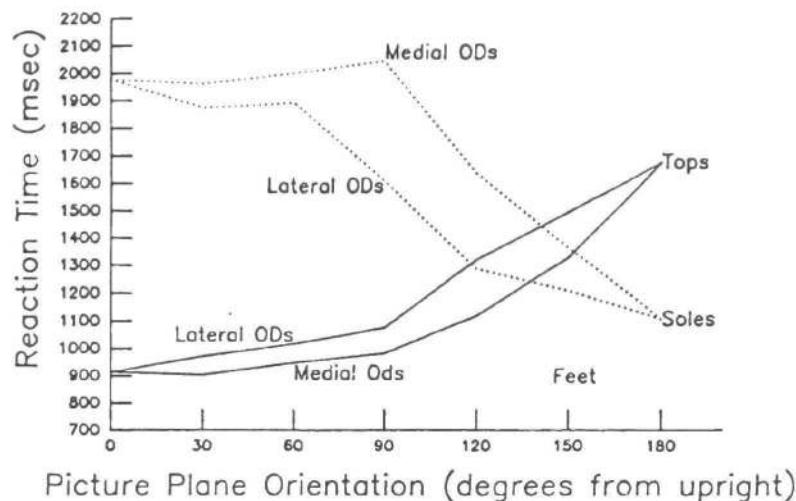


Figure 4. RT-OD functions for soles and tops of feet. Based on 4416 correct responses of 16 subjects.

To make these judgments, subjects reported imagining the rotation of their body, hand, or foot, into the orientation of the stimulus to compare the left-right features of the stimulus body or body part with the familiar left-right features of their own body or body part. Only for backs of bodies, backs of hands, and tops of feet stimuli at or near ODs of 0 degrees, did subjects report knowing "immediately" (i.e., without "thinking" about the orientation difference between themselves and the stimulus) whether the stimulus was of the left or right of the body. Accordingly, they produced their shortest RTs for such trials. Overall, RT depended on how the orientation at which the stimulus body or body-part was portrayed differed from the task-specific (or perhaps "canonical") orientation of the subjects' own body or body-part. The geometry implicit in subjects' strategies may be used to explain the difference observed between RT-OD functions for the views in Figure 1 of (i) backs and for fronts of bodies, (ii) backs and palms of hands, and (iii) tops and soles of feet.

The functions for backs of bodies and of hands, and tops of feet, are somewhat more extreme forms of the mental rotation function typically observed for another set of familiar objects (i.e., sets of letters or digits; Cooper and Shepard, 1973, and Hinton and Parsons, 1981). This kind of function is slightly sloped near a standard or most familiar orientation (e.g., upright) and is increasingly sloped as the remaining orientations increasingly differ from the standard. The RT-OD functions observed here are consistent with subjects' introspections about their performance. For trials with each of these stimuli, subjects reported imagining themselves or their hand or foot rotating from its present (or most familiar) position into the orientation of the stimulus. Subjects' introspections indicated that, for backs of bodies, backs of hands, and tops of feet, they imagined rotations about the same axis for the presentation of that stimulus at every OD in the experiment. For trials with backs of bodies, subjects reported imagining themselves rotating in a plane parallel to the frontal surface of their body, by tilting laterally about an axis perpendicular to the front of the trunk of the body and passing through the center of the body. For trials with backs of hands and tops of feet, they reported imagining rotations about long axis of the forearm and of the leg respectively.

With respect to the body of the observer, left hand and foot stimuli at clockwise ODs and right hand and foot stimuli at counterclockwise ODs are in medial positions. Similarly, left and right foot stimuli at oppositely directed ODs are seen by the observer as being at lateral positions. This aspect of spatial relationship between the observer and the portrayed position of the stimulus considerably affects subjects' performance in this task. RT-OD functions for all four hand and foot stimuli discussed here are reliably different for stimuli at lateral and medial positions. (These effects are discussed further below.)

The difference between functions for stimuli portraying the back of the hand and top of the foot at lateral and medial positions, appears to result from differences in the range of subjects' familiarity with appearance of these stimuli. The functions observed for backs of hands and tops of feet both have a flat or very slight slope (with subjects' most rapid RTs) for medial ODs near 0 degrees, which is more extensive than the range of rapid RTs for comparable lateral ODs. In each of these cases, the functions at lateral and medial ODs (beyond the initial flat portion of each function) have comparable slopes, and there is no indication of the use of alternative paths or axes of imagined rotation. Thus the longer mean RTs for lateral ODs seem related to the difference in the extent of this initial flat or slight slope. This difference

in familiarity is consistent with the differences in the range of normal motion of a hand or foot (Parsons, 1983, and Notes 3 and 7). This suggests that the normal range of physical motion can affect the imagined rotation of body parts, because the orientations of the hand or foot which become extremely familiar to the subject are within this range.

Though the functions for backs of bodies and of hands, and for tops of feet are relatively similar, there are striking differences between the RT-OD functions observed for fronts of bodies, palms of hands, and soles of feet. For fronts of bodies, subjects' introspections and RT-OD functions suggest that they imagined rotations about a single fixed, "shortest path" axis. Rotations about such an axis, which is unique for each OD, minimizes the degrees necessary to move an object from one orientation to another. The shortest path rotations for differences in orientation like those between subjects' position and that portrayed by the fronts of body stimulus, are always and only 180 degrees--regardless of the picture orientation of the stimulus (Parsons, 1983, and Note 4).

It is apparent from other results of Parsons (1983 and Note 4) that subjects' tendency to use shortest path rotations (as opposed to less efficient rotation paths) for imagining the rotation of their body is influenced by (i) properties of the difference between subjects' task-specific (or canonical) orientation and that at which the stimulus is depicted, and (ii) the familiarity of the perspective from which the stimulus is depicted seemed to influence subjects' tendency to imagine shortest path rotations.

For trials with palm stimuli at ODs that portrayed the palm at medial positions, subjects apparently imagined approximately shortest path rotations (like those used for fronts of bodies). This conclusion is supported both by the flat slope of the observed RT-OD function and by subjects' introspections. However, with palm stimuli at ODs that portrayed the palm at lateral positions, subjects apparently imagined rotations about two different axes. This difference in imagined rotation paths into the palm at lateral and medial positions is consistent with the range of normal motion of the hand. As readers can probably verify for themselves, efficient (and natural) physical rotations of the hand into the medial positions portrayed by palm stimuli follow a path different from (and shorter than) that for rotation into lateral positions. These differences in the possible (or comfortable) physical rotation paths are of course related to the mechanical properties of the anatomical structures underlying such motions (cf. Parsons, 1983, and Note 3).

For trials with the sole of the foot stimuli, subjects' RT-OD functions and introspections indicate that they did not imagine shortest path rotations. The imagined rotation path of the foot seems to have been chosen to be within the normal range of motion of the foot (Parsons, 1983 and Note 5).

Overall, the following conclusions are suggested by my recent experimental results (some of which are presented above). The RT of these left-right judgments depended on how the orientation portrayed by the stimulus body or body part differed from the task-specific (or canonical) position of the subjects' body or body part. To make these judgments, subjects spontaneously imagined a part or all of their body rotating into the orientation of the stimulus to compare the left-right features of the stimulus body or body part with the familiar left-right features of their own body or body part. In some cases, subjects imagined rotations about an axis that simultaneously corrected for all differences between the orientation of their body or body part and that of the stimulus body. Use of such a "shortest path" axis minimizes the degrees of rotation necessary to align two objects. In other cases, subjects apparently imagined rotations about two axes. They may have used this latter strategy on trials with bodies, because of difficulty finding the shortest path axis, which is unique for each orientation difference (OD). Other results of Parsons (1983 and Note 4) suggest that the tendency to imagine shortest path rotations of one's body is influenced by various properties of the OD between oneself and a goal orientation, and by familiarity with an object's appearance throughout the range of its possible rotations.

On trials with hand and foot stimuli, subjects apparently imagine very efficient--or approximately shortest path--rotations only when the path from the task-specific (or canonical) position of the hand or foot into the position portrayed by the stimulus is within the normal range of movement of the hand and foot respectively.

These conclusions seem to require a theoretical framework in which there are spatial transformational procedures or schemas specific to different stimulus domains, and rather different from one domain to another. Further, in some domains such procedures seem closely related to--indeed, seem to simulate--the actual physical performance of a spatial transformation.



## Reference Notes

1. Parsons, L.M. and Hinton, G.E. Egocentric and environmental frames of reference in the discrimination of mirror-image shapes. In preparation.
2. Parsons, L.M. Use of a "human body analogy" in the visual discrimination of abstract mirror-reflected shapes. Manuscript submitted for publication, 1983.
3. Parsons, L.M. Imagined spatial transformations in the visual discrimination of left and right hands. Manuscript submitted for publication, 1983.
4. Parsons, L.M. Imagined spatial transformations in the visual discrimination of left and right parts of the human body. Manuscript submitted for publication, 1983.
5. Parsons, L.M. Imagined spatial transformations in the visual discrimination of left and right feet. Manuscript submitted for publication, 1983.
6. Parsons, L.M. Geometrical properties of imagined rotation of abstract mirror-reflected objects. In preparation.
7. Parsons, L.M. The selection of paths for the imagined and physical rotation of an object from one orientation to another. In preparation.

## 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. Pp. 75-176.
- Cooper, L.A. and Shepard, R.N. Mental transformations in the identification of left and right hands. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, 104, 48-56.
- Cooper, L.A. and Shepard, R.N. Transformations on representations of objects in space. In E.C. Carterette and M.P. Friedman (Eds.) *Handbook of Perception, Vol. VIII*. New York: Academic Press, 1978. Pp. 105-146.
- Hinton, G.E. and Parsons, L.M. Frames of reference and mental imagery. In A. Baddeley and J. Long (Eds.) *Attention and Performance IX*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1981.
- Parsons, L.M. Mental rotation paths in the visual discrimination of left and right parts of the human body. Unpublished doctoral dissertation, University of California, San Diego, 1983.



## THE STRUCTURE OF VISUAL CONCEPTS

Joseph Psotka  
Computer-Based Systems Development  
Army Research Institute  
Alexandria, Virginia.

The research reported here addresses the question: What is the visual equivalent of a word, idea, or higher-order concept? This research is a beginning attempt to analyze visual concepts in ways that are most appropriate for future richer computational environments that will be able to process images quickly and in great numbers: massively parallel machines like those envisioned by Feldman and Ballard (1982).

Knowledge representation in artificial intelligence applications and other computer-based systems today uses many structuring schemes: feature lists, vectors, trees, networks, and production systems. In each, entities or nodes are filled with concepts representing words or lists of words. Most of these systems have developed out of a linguistic background, and reflect the symbolic processing limitations of present-day computers. This context has made images and pictures difficult to use, but it is changing.

The foundation for this research lies in the many experiments that demonstrate the great fidelity of the human pattern recognition system and its enormous capacity to store and process large numbers of visual representations in very short periods. For instance, recognizing the face of an acquaintance one has not seen for many years implies the existence of an ability to select one alternative from thousands (perhaps millions, given the many transformations a face may undergo) within a few moments. Computer-based systems cannot begin to rival this real-time feat, in part perhaps because no knowledge representation schemes appropriate to specifically visual concepts have yet been developed.

Several experiments have been conducted (and one will be demonstrated to the audience) that provide some support for the theoretical proposal that Galton's system of composite portraiture provides one model for the structure of visual concepts. This proposal also has practical implications and applications to computing. Galton used a photographic process to add several photographs of human faces onto the same picture, one on top of the other. The resulting single face may be a prototype for this particular set of faces (see Osherson and Smith, 1981 for a discussion of prototypes). His technique was used in these experiments and has been extended by Weil (1982) to a computer-controlled optical videodisc system funded by DARPA. Large scale parallel optical computers offer the possibility of dramatic future enhancements.

The experiments in this report are an impoverished beginning for the very large task ahead. They have established that:

- 1) Composite faces can be recognized after viewing the set of component faces they comprise;
  - 2) The similarity of components to composites correlates with recognition performance;
  - and 3) Composites are more attractive than their component faces.
- These findings are hardly earth-shaking in their own right, but they fit well with the general framework that composites are prototypes. Basically they provide an incentive for looking for more critical facts that might flesh out the critical features of visual concepts.

Galton's technique of superimposing whole faces in registration to develop an average face provides a wholistic procedure for creating prototypes and making them visible. The first question that the procedure raises is: can people recognize these wholistic composite faces after viewing their components? Second, do these prototypes accurately represent critical features and relations among the components they comprise? Third, is there anything unique or distinct about these prototypes, not found in their components?

Finally, the Galtonian process of composite portraiture provides a reduced analogue for convolution and crosscorrelation processes (as used in holography) that have provided the impetus for models of memory (cf. Psotka, 1977; Metcalfe and Murdock, 1981). Are these composites sensitive to common memory research manipulations?

**SUBJECTS:** Four groups of 25 undergraduates were used in the first half of the experiment. A fifth group of 20 undergraduates were used in the second half. All students were naive about the purpose of the experiment and participated as part of a course requirement at the University of Waterloo.

**PROCEDURE:** 100 full-face photographs were selected randomly from male yearbooks with the restriction that there should be no facial hair or eyeglasses. These faces were randomly grouped into ten lists of ten faces whose composites were photographed.

The first four groups were shown only one-half of the 100 faces in five lists of ten faces, each list followed by a forced choice pair of composites (See figure 1). They were asked to pick the more familiar of the two composites. Only one of the composites was composed of the list of ten faces previously shown. In order to ensure that each student examined each face carefully, they were asked to rate the attractiveness of each face on a ten point scale. The faces were presented on slides shown in groups with order of presentation of the lists counterbalanced.

The fifth group rated the similarity of the paired composites to each of the two sets of ten components on a ten point scale. This task was performed individually using prints of the slides.



**RESULTS:** Recognition performance showed that people were able to recognize a composite on the basis of viewing the prototypes. Percent correct ranged from 48 (chance) to 92. Eighteen out of 20 groups were above chance (sign test,  $p < .01$ ). There was an effect of order of presentation. The first prototypes shown were recognized best.

The similarity ratings showed that a component face was more similar to the composite into which it entered than to the distractor composite ( $t=5.23$ ;  $df=99$ ;  $p < .01$ ). A significant correlation was found between the similarity ratings and percent correct recognition ( $r=.86$ ;  $df=8$ ;  $p < .01$ ).

The composites were each judged much more attractive than the component faces. All ten lists showed this effect. Only six of the 100 faces were judged more attractive than the least attractive composites. These faces were also the most similar to the composites.

**CONCLUSIONS:** The evidence is clear that people can recognize composites on the basis of their memory of previously viewed components. Of course, they could use distinctive features from each unique face stored in memory, or they could be creating a composite as they are viewing the set of faces. On this point, the evidence is not decisive; but then this is a venerable dispute that traces back to Locke and Berkeley. I would hardly expect one experiment to provide a resolution. Nevertheless, I believe that something like these composites is being formed in the visual/memorial system.

The high correlation between the difference in similarity ratings and percent correct recognition provides evidence that people base their recognition process on the entire group of faces they have seen. If the group of faces is remembered as a whole, then in some sense it can be represented best by a composite that is an average of the entire group. It seems impossible to distinguish between a set of separate and unique individuals and an amalgam --- schema or prototype --- that acts separately unless an emergent property of the amalgam can be found. This emergent property may be attractiveness. There seems to be no satisfying explanation for the increased attractiveness of the composites except that they are more representative of the whole set of faces (beyond this experiment) encountered in our experience. They are more like the prototypical face we have unconsciously created as a standard in its physical rather than personality characteristics.

In summary, there are four main pieces of evidence to leave with: First, there exist longstanding conceptions and theories with elegance and considerable power. They address the existence of global concepts as unique ideas or associations, separate from individual stimulus items, events, or episodes in memory. Second, the techniques reported here have uncovered several facts that may not provide critical support for these theories, but lend credence to them and offer avenues for research to discover facts that would distinguish them from competing theories.

Third, the findings are of some interest in their own right. The relations among motivation and cognition are still wide open to be explored by cognitive scientists for they have hardly advanced since Wolfgang Kohler's monumental effort in "The Place of Value in a World of Fact".

Finally, these techniques applied with greater computational power may be useful tools for other purposes: to create beautiful images, or for criminal identification as Weil (1982) began to explore.

#### REFERENCES

- Feldman, J.A. and Ballard, D. H. Connectionist models and their properties. *Cognitive Science*, 1982, 6, 205-254.
- Metcalfe, J. and Murdock, B. B. An encoding and retrieval model of single-trial free recall. *JVLVB*, 1981, 20, 161-189.
- Osherson, D. N. and Smith, E. E. On the adequacy of prototype theory as a theory of concepts. *Cognition*, 1981, 9, 35-58.
- Psotka, J. Syntely: Paradigm for an inductive psychology of memory, perception, and thinking. *Memory & Cognition*, 1977, 5, 553-560.
- Weil, P. About Face: Computergraphic synthesis and manipulation of facial imagery. MIT, 1982, MS dissertation.

# **Paper Session #3**





# THE ROLE OF INTERPERSONAL GAMES IN PERLOCUTIONARY ACTS

Gabriella Airenti\*, Bruno G. Bara\*, Marco Colombetti\*\*

\*Unità di ricerca di intelligenza artificiale, Università di Milano

\*\*Progetto di intelligenza artificiale, Politecnico di Milano

## Abstract

The analysis of perlocutionary acts is a fundamental component for a theory of human communication which accounts for the response of interacting subjects. We introduce two knowledge structures, namely games and contracts, which are used by the mental processes underlying perlocutionary acts. We present our model through the discussion of an example of game; in particular we focus on the problem of shifting from the wants of the actor to the wants of the partner.

## 1. Introduction

Cohen and Perrault (1979) and Perrault and Allen (1980) propose a formal model of illocutionary acts, based on speech act theory as developed by Searle (1969). Starting from these results, we suggest a model of perlocutionary acts in order to explain: (i) how an actor A plans speech acts to obtain a specific response from a partner P, and (ii) how P decides to produce a response. Our interest here is with point (i).

Following Cohen and Perrault (1979) we assume that:

- A performs an illocutionary act whose effect is P's recognition of A's want that P performs a specific action  $\pi$ ;
- a process called CAUSE-TO-WANT is triggered by the illocutionary effect and generates the perlocutionary effect that P wants to perform  $\pi$ .

The main point here is that the process CAUSE-TO-WANT actuates the transition from A's wants to P's wants. When both A and P are human systems, this transition deserves a further investigation. We claim that within a cognitive system a want can only be generated:

- by a knowledge structure of the kind of Schank and Abelson's life themes (1977);
- as a subwant of a previously existing want.

In the following we shall introduce two knowledge structures, namely games and contracts, as useful tools for a theory of human communication which provides for an adequate treatment of want generation.

## 2. Games in perlocution

The mental processes underlying perlocutionary acts make use of knowledge structures that we call contracts and games.

A contract is a cluster of actions involving two actors. When the contract is activated the two actors mutually assume the obligation to perform their roles (see Airenti, Bara and Colombetti, 1983).

A game describes the interactions of two actors (the players), as regulated by scripts within a specific context. The use of a game for planning perlocutionary acts will be illustrated by the following example, in which we reconstruct the process of plan formation of an actor A. Let us suppose that A wants to be driven to a location LC by a partner P; to fix the context, let us assume that A is a guest in P's house.

To reach his goal A must cause P to want to perform his role. We reduce such a CAUSE-TO-WANT action to: (i) three inferences, made possible by a script, a game and a theme; (ii) a CONVINCING action (see Fig. 1, where a slot-filler formalism has been adopted: symbols in capitals are constants, small single letters are variables).

The first inference is based on the script GIVE-A-LIFT; it leads to assume the driver's want, PARTICIPATE-IN-SCRIPT, as a condition of the driver's want DRIVE-VEHICLE. The use of a script in this kind of inference is that if an actor wants to participate in the script, then he wants to perform the action assigned to him in the script. The second inference is based on the game A-GUEST-P-HOST; it leads to assume P's want to PARTICIPATE-IN-GAME as a condition of P's want to PARTICIPATE-IN-SCRIPT. The introduction of the concept of game has the aim of providing a motivation for an actor to participate in a script within a specific context. Contrary to scripts, games are defined for two particular players. We assume a principle of interaction which represents the subjects' need of participating in interpersonal games. This principle is realized through the theme INTERACTION, which allows A to draw the third inference. This leads to assume A's want to PARTICIPATE-IN-GAME as a condition of P's want to do the same. The major point here is the shift from the actor's want to the partner's want.

The three inference steps lead to a formula which is the effect of a CONVINCING perlocutionary act, detailed in Fig. 2. We reduce the CONVINCING action to an inference and to the first action of the script GIVE-A-LIFT. The inference is made possible again by the game A-GUEST-P-HOST. Its meaning is that an actor is supposed to participate in a game if he manifests his intention to participate in a script belonging to that game, within its applicability context. Eventually, A's intention to participate in the script can be deduced from the fact that A performs the first action mentioned in the script (in this case the request). This amounts to assume that the execution of the first action counts for an opening move of the corresponding game.

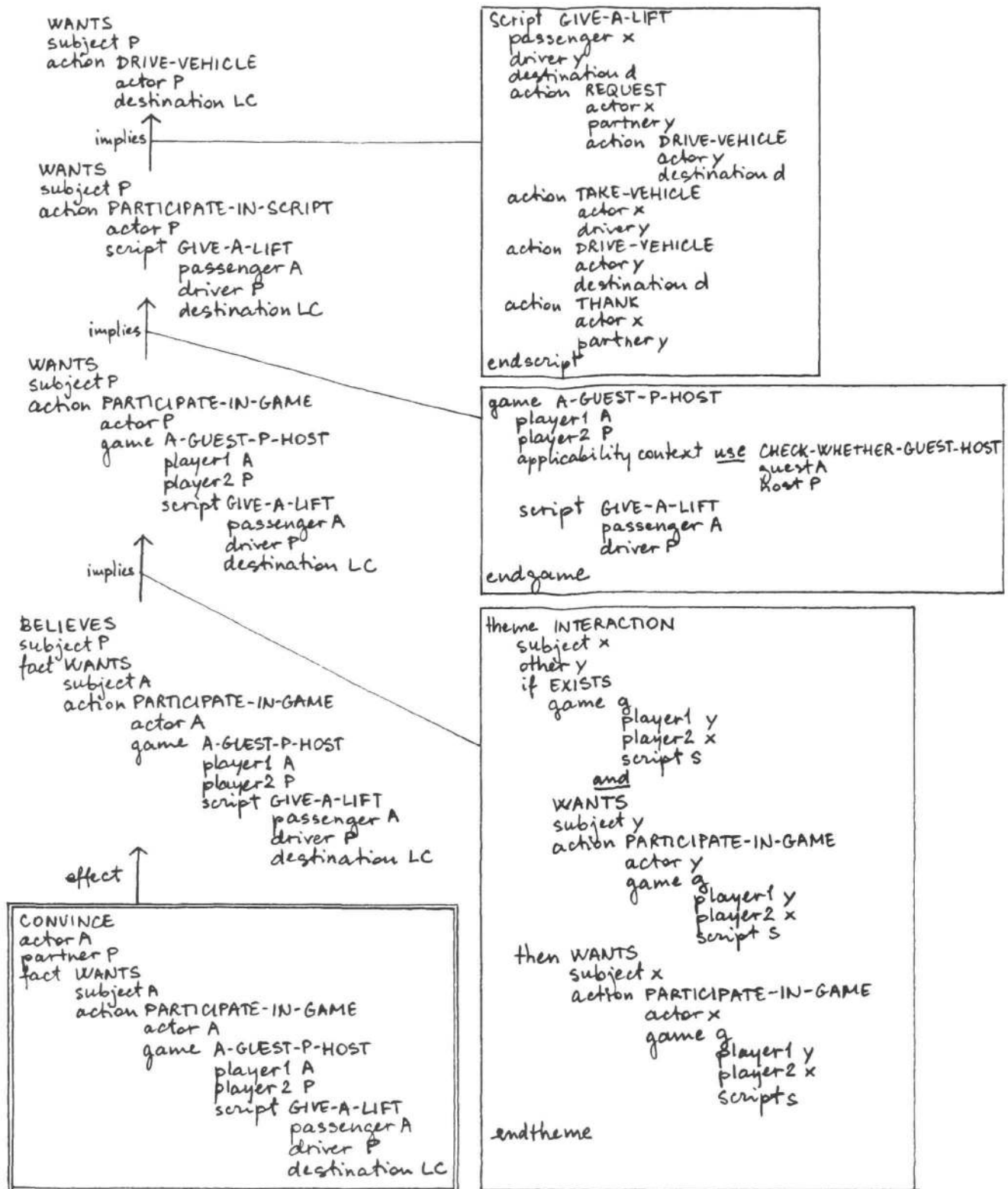


Fig1. The analysis of the CAUSE-TO-WANT action

The given analysis models the process used by A to plan a perlocutionary act through a game. The same result could be obtained via a shortcut by assuming that A's request directly generates P's want to perform the DRIVE-VEHICLE action. The full process is however necessary to account for:

- deceit: an actor may make the first move of the script in order to convince a partner to play his role, without having the intention to play his own role fully;
- failure recovery: when A executes his plan, a failure may occur at any point; the complete plan allows A to recover from the failure by partial replanning.

To give an example of failure recovery, let us suppose that the actor A receives no answer to his request; by using only the shortcut, A would have no possibility but giving up or repeating the same request. On the contrary, through the analysis of the complete plan given in Figs. 1 and 2, A is able to make a guess about the failure point, and to partially replan. Possible failures could occur in connection with any of the knowledge structures involved, namely the script, the game, the theme. For instance, A may assume that the script GIVE-A-LIFT is not part of the game A-GUEST-P-HOST for his partner P. In this case, A replans his perlocutionary act trying to replace the script GIVE-A-LIFT with another script associated to the same game or to a different one.

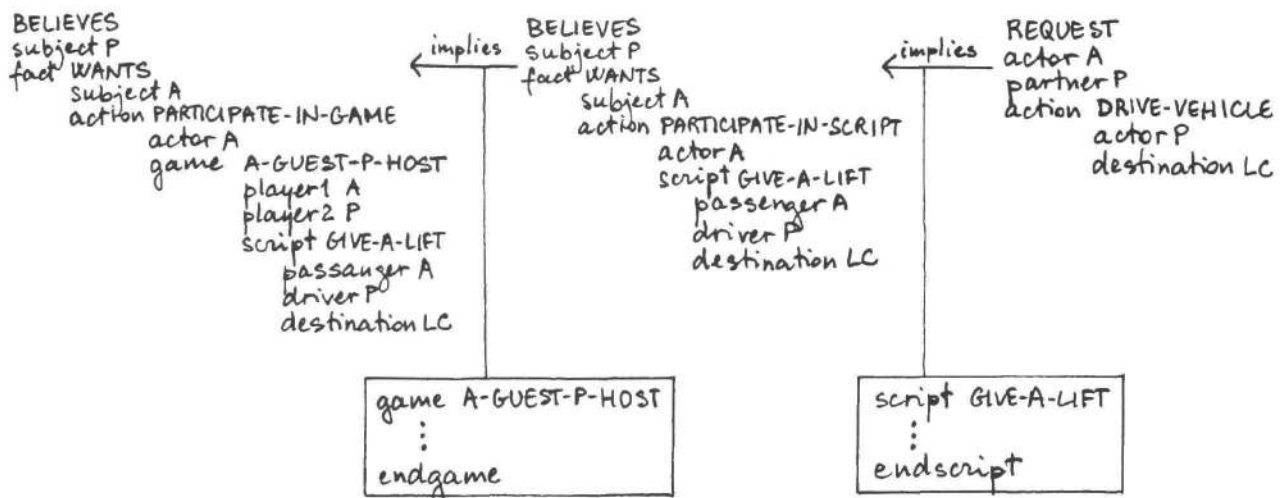


Fig.2. The analysis of the CONVINCE action

### 3. Discussion

To explain how an actor plans speech acts to obtain a response from a partner is a relevant question for cognitive modelling. The model we present could serve as a basis for an artificial system able to plan perlocutionary acts taking into account possible failure recoveries and deceptions.

Moreover, our proposal seems to provide a first psychologically adequate explanation of the cognitive processes necessarily involved in the production of perlocutionary acts. In particular, we believe that a knowledge structure like the game is needed to explain the transition from an actor's want that a partner performs an action  $\pi$  to the partner's want to perform  $\pi$ . Such a transition has been reconducted to a basic principle of interaction, which is assumed to provide for the motivation of action.

### References

- Airenti G., Bara B.G., Colombetti M., 1983. Planning perlocutionary acts, Proceedings 8th IJCAI, Karlsruhe
- Cohen P.R., Perrault C.R., 1979. Elements of a plan based theory of speech acts, Cognitive Science, 3, 3
- Perrault C.R., Allen J.F., 1980. A plan based analysis of indirect speech acts, American Journal of Computational Linguistics, 6, 3-4
- Schank R.C., Abelson R.P., 1977. Scripts, plans, goals and understanding, Lawrence Erlbaum Press, Hillsdale, N.J.
- Searle J.R., 1969. Speech acts, Cambridge University Press, London



# NINETY YEARS OF MENTAL METAPHORS<sup>1</sup>

Dedre Gentner  
Bolt Beranek and Newman Inc.  
Cambridge, Mass.

Jonathan Grudin  
MRC Applied Psychology Unit  
Cambridge, England

## Abstract

In this paper we seek to trace the way in which psychologists's concepts of the mind have evolved over the roughly ninety years since the study of empirical psychology began in America. We examined metaphors used by psychologists to describe mental phenomena, based on a corpus of mental metaphors used in the journal Psychology Review from 1894 to the present.

The chief finding was that the nature of the mental metaphors changed over time. Spatial metaphors and animate-being metaphors predominate in the early stages, declining later in favor of systems metaphors from mathematics, the physical sciences and artificial intelligence. A secondary finding was that the numbers of mental metaphors varied: They are more prevalent in the early and late stages of the century than in the mid-stages. These patterns are interpreted in terms of the evolution of psychologists' models of the mind.

## Ninety Years of Mental Metaphors

In this paper we examine historical changes in the metaphors used by American psychologists to describe mental processes. Our aim is to use changes in metaphoric language to trace changes in the models psychologists have held of the mind. Three assumptions underlie this work. First, we assume that researchers bring to their field of study a theoretical framework -- which may be more or less explicit -- in terms of which they construe the phenomena they observe. Second, we assume that these frameworks can change over time, sometimes quite rapidly (Kuhn, 1962). Third, we assume that analogies and metaphors are used in the invention and organization of ideas in science (Gentner & Gentner, 1983; Hesse, 1966).

Cognitive psychology during the past hundred years seems a prime example of a field in which conceptual change has been rapid and extensive. It would be useful to have a method for tracing changes in the Zeitgeist. One way to do this might be by examining the metaphors used by psychologists. It has been argued persuasively that metaphors from other domains have played a role in the shaping of psychological theory. Roediger (1980) noted several distinct metaphors for human memory, ranging from Freud's rooms-of-a-house model to Atkinson and Shiffrin's storage-box model.

If indeed the metaphors used in psychology reflect the way that researchers have conceived of the domain, then changes in the kinds of metaphors used to describe the mind may provide an unobtrusive measure of changes in the conceptual paradigms used in American psychology. With this in mind, we undertook to collect a representative sample of metaphors of the mind. We chose as our source the journal Psychological Review, since it has a history of broad representation of major work in psychology that dates back to 1894. Thus, our project was (1) to sample Psychology Review

---

<sup>1</sup>Support for the preparation of the paper was provided in part by the National Institute of Education under Contracts No. HEW-NIE-C-400-80-0031 and HEW-NIE-C-400-81-0030.



systematically from 1894 to the present for mental metaphors; (2) to classify the resulting metaphors according to their base domain (their analogical domain, or domain of origin); and (3) to note any changes in the numbers or types of metaphors used across time.

### The Survey

The volumes surveyed spanned nine decades, beginning with 1894, then 1905, 1915, and so on through 1975. We examined all articles in the first issue of each volume. All mental metaphors were recorded on their first occurrence in a given article. A mental metaphor was defined as a comparison in which either the mind as a whole or some particular aspect of the mind -- ideas, processes, etc. -- is or explained in terms of a nonmental domain. We included everything that seemed a possible metaphor, including many frozen or conventionalized metaphors, such as "mental health" or "intellectual level." In each article, only one instance of any given metaphor was collected. However, when several terms occurred in an extended metaphor, all of the terms were collected. An example of such a system occurs in James (1905). The phrases "an idea encountering a resisting idea," "an idea moving under its own momentum," and "ideas overcoming an obstacle" were each recorded, although they are clearly part of the same extended metaphor.

Categories of metaphors. Out of a total of 68 articles, 48 contained mental metaphors. We found a total of 265 metaphors for mental phenomena. After the set of metaphors was assembled, we sorted them into categories drawn from a common domain. The sorting was done by the content of the metaphors, crosscutting decade of origin. Where more than one category or subcategory might apply, we used the most central and specific features of the metaphor to select among alternatives. Our sorting yielded 20 subcategories, which combined into four major categories of metaphor: Animate-being metaphors (23 instances), Neural metaphors (16 instances), Spatial metaphors (61 instances), and Systems metaphors (80 instances).

In Animate-being metaphors, ideas or aspects of the mind are likened to creatures; e.g., "Through lying, the mind grows wary or strong from swimming against the stream." (Dewey, 1904). In Neural metaphors, the analogical domain is some version of the physical nervous system, as when it is stated that word meanings are stored as mental images "located in different regions of the gray cortex of the brain, and joined together in a unit by a series of association-tracts which pass in the white matter under the cortex" (Starr, 1894). or movement of objects in space, as in "things active against a background of consciousness" (Strong, 1894). Systems metaphors are those that liken some mental phenomenon to a system of lawfully constrained interactions among elements. Often, they draw on a physical or mathematical system or on an artificial device as their analogical domain: e.g., "fusion of ideas" (Peterson, 1935), or "critical band behaving like a variable band-pass filter" (Zwicker & Scharf, 1965) Instances from each of these categories are given in Table 1 for early, middle and recent periods.

In addition to the four major categories, there were two other categories: "Conventional" metaphors (71 instances) possible metaphors whose metaphoric associations seem to have been lost: e.g., "mental health" and "intellectual growth"; and "Idiosyncratic" or unclassifiable metaphors (14 instances).

Patterns of distribution. The major finding is a shift in the categories of metaphoric domains used over time. Figure 1 shows the number of metaphors used in each of the four major categories in each of the three tri-decade blocks. In the early samples, Spatial metaphors and Animate-being metaphors dominate. There is a sharp drop across time in the number of Animate-being metaphors, along with a less severe reduction in the use of Spatial metaphors. Systems metaphors show the opposite trend: Starting as an unimportant category, with 5 members in the first tri-decade, System metaphors gradually come to predominate. Statistical analyses indicate that

Table 1

## Examples of the Four Major Categories of Metaphor by Tridecade.

Early: 1894-1915	Middle: 1925-1945	Recent: 1955-1975
<b>ANIMATE</b>		
10) Through lying, mind grows wary or strong from swimming against stream.	34) Reaction arcs block each other, varying in tension, until one waxes strong enough.	85) Super discriminating pre-perceiver who selectively prevents recognition.
11) Ideas struggle with one another.		85) Ego defenses.
<b>NEURAL</b>		
5) Associations among images like white matter connecting regions of gray matter.	34) Thinking is neural impulses shifting along associative fibers from one area to another	81) Inhibitory processes.
11) Wider ideas shortcircuit smaller ideas.	63) Anger shortcircuits excitation into the parasympathetic system.	81) Loudness perhaps proper- to number of mental impulses
<b>SPATIAL</b>		
11) Anything hiding in the back-ground is not mental activity.	41) Habitual connections between ideas.	81) Critical Band is formed.
21) Tracing is to a photograph as memory is to immediate attention.	63) Fear inundates the sympathetic nervous system.	82) Reservoir model for Fixed Action Pattern.
<b>SYSTEMS</b>		
11) A body moves in empty space by its own momentum as when our thoughts wander at their own sweet will.	48) Nervous system is like a switchboard mechanism.	72) $O = R/R + I$ , Where R = # of relevant elements I = # of irrelevant elem. O = conditioning constant
21) Associative force	49) Goal gradient: positive/negative transfer.	94) Serial iterative operations

Animate-being, Spatial and Idiosyncratic metaphors decrease significantly in numbers. Systems metaphors increase across time. Neural metaphors and Conventional metaphors remain constant in number.

Variation in overall metaphor usage. A secondary finding is a U-shaped variation in the overall numbers of metaphors used in different periods. Metaphors for the mind are abundant at the outset of our sample (1894-1915), drop sharply from approximately 1925 to 1945, and rise to even greater numbers during the most recent tri-decade (1955-1975). The dip in mental metaphors during the middle third of our survey (1925-1945) seems part of a general decline in the use of mentalistic language due to the influence of behaviorism. Articles from this period tended to be straightforward reports of data, devoid of any discussion of the internal workings of the mind.

### Conclusions

The most interesting finding is the shift in the kinds of domains from which metaphors were drawn. Why did systems metaphors replace the animate and spatial metaphors that predominated in the early stages? We turn now to consideration of the explanatory goals these metaphors were intended to serve. To begin with, we pose three questions that will serve to organize the discussion: (1) What is the function of metaphor in scientific explanation; (2) Are some explanatory metaphors better than others; and finally, (3) If so, have the mental metaphors in psychology improved over time?

In scientific exposition, an analogy can allow prediction by mapping known relationships from a familiar domain into an unknown target domain (Gentner, 1980; Gentner & Gentner, 1983). The predictive usefulness of a metaphor reflects not only the precision and plausibility of its correspondances, but also its systematicity: the degree to which its inferred predicates form a mutually constraining system. Systematicity is valued in scientific explication, because interrelations among the inferred predicates allow new predictions.

Have psychology metaphors become more systematic? A remark by William James (1890) suggests this possibility: "At a certain stage in the development of every science a degree of vagueness is what best consists with fertility." James and other earlier writers may have used metaphor in an expansive, less precise manner. Certainly some of the early animate metaphors seem to lack systematicity; for example, "Memory moves more easily from a name to a person [its referent] than the reverse, as a fish swims more easily from upriver down to the ocean." (Starr, 1894). In contrast, when the analogical domain is a mathematical or physical system, concatenations of immediate predictions into further predictions are possible. Algebraic metaphors such as the learning theory equation ( $\theta = r/r+i$ ) are one example (Restle, 1955). The systematic nature of the analogical domain allows a set of interrelated predictions: e.g., that  $\theta$  (the conditioning constant) should rise with the number of relevant elements ( $r$ ) and decrease with the number of irrelevant elements ( $i$ ); that the ratio of relevant to irrelevant elements should be  $\theta/1-\theta$ , and so on. The move towards systematic analogies was surely partly motivated by desire for this kind of predictive power.

Now we turn to the specific question of why, in recent times, computer systems metaphors have dominated over other systems metaphors. Certainly the adoption of these metaphors does not guarantee either rigor of application or interestingness of results. Use of the computer metaphor does not even guarantee avoidance of animism. Terms like "retrieving", "detecting" and "searching" can all describe human behaviors as well as machine operations, and this ambiguity is sometimes exploited in vague analogizing. It has been observed that an entire homunculus can be concealed within one processing box in a flow diagram (Mandler, 1978). Nevertheless, a computer analogy can represent a genuine simplification, if the powers of the individual processors are sternly limited. As Dennett (1978) puts it, "If one can get a team or committee of relatively ignorant, narrow-minded, blind homunculi to produce the intelligent behavior of the whole, this is progress."

It is tempting to conclude that there has indeed been a change in the "degree of vagueness" tolerated in modelling, and that the current analogies are more conducive to progress in understanding the mind. But, according to the thesis assumed here, our judgements must be cautious, since we see through the metaphors of our time. Our own frameworks remain to be evaluated.

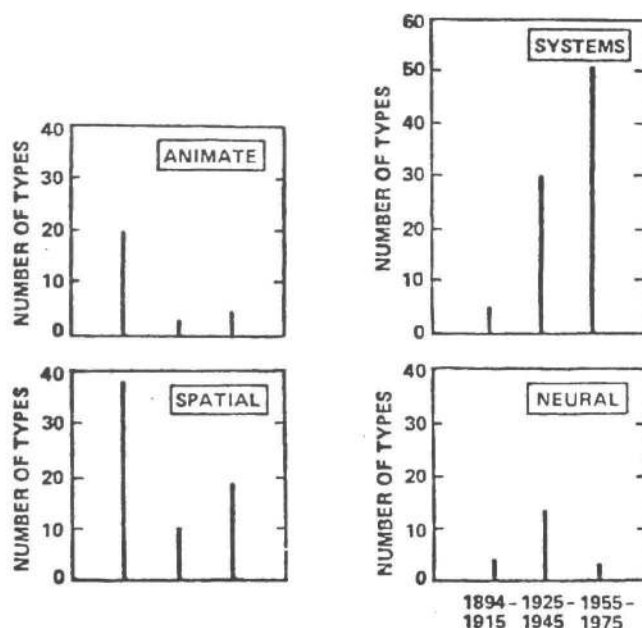


Figure 1. Numbers of metaphors in each tri-decade for each of the four major categories.

## References<sup>2</sup>

- Dennett, D.C. Brainstorms. Montgomery, Vermont: Bradford Books, 1978.
- Gentner, D. & Gentner, D. Flowing waters or teeming crowds: Mental models of electronic circuits. In D. Gentner & A. Stevens (Eds.), Mental models. Hillsdale, N.J.: Erlbaum, 1983.
- Gentner, D. Structure-mapping: A theoretical framework for analogy. Cognitive Science, in press.
- Hesse, M. Models and analogies in science. South Bend, Indiana: University of Notre Dame Press, 1966.
- James, W. The principles of psychology. New York: Holt, 1890.
- Kuhn, T. The structure of scientific revolutions. Chicago: University of Chicago Press, 1962.
- Lakoff, G. & Johnson, M. Metaphors we live by. Chicago: University of Chicago Press, 1980.
- Mandler, G.A., An ancient conundrum: Review of K.R. Popper and J.C. Eccles, The self and its brain, Science, 1978, 200.
- Roediger, H.L. Memory metaphors in cognitive psychology. Memory and Cognition, 1980, 8, 231-246.

---

<sup>2</sup>To save space, references to surveyed articles from Psychological Review are omitted: they can be found in the first issue of their year of publication.





# The Evaluation of Cognitive Constructs Using Structural Equation Modeling

Morton Ann Gernsbacher & H. H. Goldsmith  
University of Texas at Austin

Though, in anthropomorphic terms, the field of cognitive psychology has barely grown out of young adulthood, its development thus far has been markedly influenced by several of its sister sciences (e.g., theoretical linguistics, artificial intelligence, engineering science, philosophy). These influences are most noticeable within the structural models cognitive psychologists propose to describe various aspects of cognitive processes. This paper presents an approach to structural modeling that arose from work within the field of econometrics and has recently been successfully applied to describe behavioral processes within the fields of sociology and biometrical genetics. We believe this approach to be particularly well suited for modeling cognitive behavior because it helps to resolve several of the paradoxes involved with other methods of modeling.

Rather ironically, cognitive psychologists demonstrate less catholic tastes in the methodology they employ to verify their models than they do in outlining their models. Most cognitive psychologists collect behavioral data from classical 2 by 2 designs and then analyze these data with standard analyses of variance. Some potential problems arise when we attempt to extrapolate from two or four cells of means (or a series of such sets of data) to a more elaborate description of an underlying process. First, although we manipulate our experimental variables in a binary fashion, very often these variables occur naturally in a continuous distribution and it is this latter distribution we often imply in our models. Second, in order to obtain laboratory control, we systematically investigate the effects of different variables by pitting them in a series of one-to-one contrasts. Thus, the entire model is seldom tested simultaneously and we never gain an appreciation for the extent to which the model can account for all the phenomena it purports to explain. Third, the degree of the relationship between the theoretical constructs represented by the model and the variables used to measure these constructs is often left unquantified. For example, how well is the construct we call "long-term memory" indexed by the percentage of words correctly recalled after a two-week interval? In actuality, any time we test a theoretical model we are simultaneously testing the adequacy of a "measurement model." With the approach we shall describe today, this fact is made explicit. Moreover, the processes of model fitting and model testing are integrated into the same procedure.

## Overview of Structural Equation Modeling

Structural equation modeling (SEM) is a comprehensive system for testing systems of linear hypotheses involving both observed, or "manifest," variables and theoretical constructs, or "latent" variables. Best known among the family of SEM approaches is the Linear Structural Relationships (LISREL) approach (Joreskog & Sorbom, 1978) which we used in the present demonstration. It is important to note that the SEM approach requires that the investigator have an explicit theory guiding his/her research and that all latent constructs in the theory be assessed, preferably with multiple convergent measures. The implementation of SEM involves three major steps:

(1) Specification of the model. The SEM approach requires the investigator to specify, a priori, a model in which theoretical constructs are hypothesized to be functionally related to observed variables. In cognitive psychology, these variables might be behavioral measures such as reaction time or performance accuracy on a laboratory task or they could be stimulus characteristics such as orthographic regularity or semantic meaningfulness as indexed by a normative scale. The theoretical constructs are such entities as lexical familiarity, memory span, or the structure of semantic categories. Furthermore, "causal" (or functional) relationships among the theoretical constructs must be specified. Thus, SEM can be thought of as a procedure that combines elements of traditional multiple regression, factor analysis, and path analysis.

(2) Model estimation. Once the model is specified, the values for the parameters in the model are simultaneously estimated using an iterative procedure, which in the case of LISREL is a maximum likelihood algorithm. The input for this analysis is a variance-covariance matrix for all of the observed variables measured in the study. The magnitude of three types of parameters can be estimated. Values for the hypothesized causal relationships can be interpreted as partial regression coefficients. Values for relationships that are not specified as being causal can be thought of as correlations. Finally, the residual, or unexplained variation, in the latent variables and in the manifest variables is estimated. The statistical significance of each of these parameters (with exceptions noted later) can also be computed.

(3) Evaluation of goodness-of-fit of the model. The statistical evaluation of the overall fit of the model to the data is a crucial element of the SEM approach, an advantage that distinguishes SEM from most other data analytic procedures. A chi-square statistic, with its associated degrees of freedom, is the most common indicator of the likelihood that the observed variance/covariance matrix could have emerged if the specified model were "true." The number of degrees of freedom in a model is the remainder when the number of parameters being estimated is subtracted from the unique number of observed variances and covariances. If this difference is negative, the model is, of course, not identified. Larger values of chi square, relative to the degrees of freedom, indicate a poorer fit of the model to the data. It is this goodness-of-fit evaluation that places SEM within the group of confirmatory, rather than exploratory, statistical procedures.

#### The Constructs of Category Structure and Category Verification

Over a decade of research in the field of cognitive psychology has been aimed toward investigating human semantic memory. The general consensus emerging from this body of work is that semantic memory is organized in a highly systematic and orderly fashion. One of the most commonly described units of organization within this store is the semantic category. Several principles have been proposed to underlie the organizational structure of such categories. The most popular of these are the principles of association frequency, semantic distance, featural overlap and typicality.

According to the principle of association frequency, membership in a semantic category is a function of the frequency with which a category member, such as ROBIN has been previously associated with a particular category concept, such as BIRD, and vice-versa. Many of you will recognize this principle as underlying many specimens of the very familiar breed of network models of conceptual knowledge. When speaking about semantic categories, we have simply substituted the term "category concept" for the term "superordinate" and the term "category member" for "subordinate." The principle of association frequency is usually assessed by collecting normative data upon the frequency with which subjects will mention a category member in response to a category name, and vice-versa.

A second principle proposed to underlie the structure of semantic categories is based purely upon degree of intra-category similarity. The general procedure used to assess this principle is to ask subjects to rate the similarity of pairs of members from a particular category. These data are then submitted to a multidimensional scaling procedure that places the category members in Euclidean space such that the metric distances between category members is inversely and monotonically related to their semantic similarity. And, according to this principle, category membership is a function of a member's scaled position within the multidimensional configuration revealed for that category.

A third principle that has been proposed to underlie the structure of semantic categories is that of typicality. Typicality simply refers to the degree to which each member of a category is believed to be a good exemplar of its category. For example, most subjects will rate a ROBIN as being a very typical member of the category BIRDS while a CHICKEN is usually rated to be much less so.

A fourth principle that has been proposed to underlie the structure of semantic categories involves the notion of features. Features are attributes or properties of a



semantic concept. Though the possession of a feature by a concept can only be present or not present, semantic features themselves are believed to vary in a more continuous fashion in regard to their importance in defining category membership. For example, a feature of the category BIRDS such as "has feathers," might be more important in defining category membership, while a feature such as "perches in trees" might be less criterial. Given that each category member can also be described by a set of characteristic features, according to the principle of featural overlap, membership in a semantic category is a function of the number and type (more or less criterial) of categorically descriptive features shared between the category member and the category concept.

Clearly, each of these principle that have been proposed to underlie the structure of semantic categories implies the existence of a theoretical construct, namely Category Structure. Thus, in the nomenclature of SEM, Category Structure is a latent variable. Also common to each of these structural principles is an empirical prediction. Each principle predicts that Category Structure affects behavioral processes. The laboratory measure commonly used to evaluate this prediction is performance on a timed category verification task. In this task, subjects are presented with the name of a category member and asked to verify that it belongs to its appropriate category. Both the speed with which a subject responds (i.e., reaction time) and his/her accuracy of response (i.e., error rate) are provided. Thus, it has been proposed that measures of these four principles are related to the latent variable Category Structure. Category Structure, in turn, has been proposed to influence another latent variable, what we will call Category Verification. The latter variable is measured by reaction time (RT) and error rate. Here is a prime example of the proposal of implicit constructs and their causal relationships that requires evaluation in a simultaneous fashion.

In this study, we collected several independent sets of data upon one rather large sample of items. These items were eight semantic categories (viz., FRUITS, VEHICLES, FURNITURE, VEGETABLES, BIRDS, SPORTS, CLOTHING) and twenty each of their respective members. The data collection was arranged into two stages. In the first stage, measures of each of the four structural principles were obtained for the entire set of items in procedures identical to those employed by previous researchers. In the second stage, measures of performance (i.e., speed and accuracy) on the speeded category verification task were obtained, using the previously measured items as experimental stimuli. A different group of 50 undergraduates at the University of Texas participated in each aspect of the data collection. These subjects provided us with a data base composed of 24,000 measures of association frequency, 4826 measures of feature criteriality, 80,000 measures of feature possession, 76,000 measures of semantic distance, 800 measures of typicality, 12,750 RTs (with the effects of word length removed), and 136 erroneous responses. These data were reduced to a more manageable 6 X 6 (four structural principles and two performance measures) correlation matrix that we proceeded to analyze using the SEM approach.

#### Application of Structural Equation Modeling

From the theoretical guidance outlined above, the model in Figure 2 has been specified, estimated, and evaluated. Before discussing the model itself, we should clarify the notation used in Figure 1. Manifest variables are represented by rectangles, latent variables are depicted as circles, the direction of causal or functional relationships is specified by arrows, and unexplained relationships are denoted by curved lines. Each figure that we could draw using the symbols in Figure 1 and the conventions of path analysis specifies a series of linear equations that are simultaneously solved by the LISREL procedure. For the analysis presented here, we used the RAM parameterization (McArdle & McDonald, 1981) of version IV of the LISREL program (Joreskog & Sorbom, 1978).

Let us now consider each of the portions of the model. On the left side are the manifest variables hypothesized to represent different, but correlated, aspects of the latent variable Category Structure. In one sense, we can think of Category Structure as a factor and the standardized partial regression coefficients .69, -.57, .82, and .79 as

factor loadings for the 4 observed variables. Note that these coefficients are high and that the residual variance in each manifest variable is low. The negative value for semantic distance is simply due to the fact that the Multidimensional Scaling program (ALSCAL) used to derive the measure scales dissimilarities rather than similarities.

The two undirected relationships at the extreme left of Figure 1 represent unexplained associations between residual variation between Semantic Distance and Typicality and between Typicality and Featural Overlap. The existence of these weak but statistically significant relations means that there is systematic covariance within these pairs of manifest variables that is not common to the other two variables.

Switching our attention to the right side of Figure 1, we find the measurement model for Category Verification. Gernsbacher (1982) has empirically demonstrated that a combination of Reaction Time and Error Rate is a much more comprehensive measure for evaluating performance in many speeded cognitive paradigms than either of the two measures alone. Even the low error rate observed in these data (viz., mean percentage of error = 5%) contributes substantially to the Category Verification latent variable. We do notice, however, that RT is the stronger contributor.

The key prediction of the model is that Category Structure bears a functional relationship to (or "causes") performance on the Category Verification task. Thus, we examine the directed path in the center of Figure 1. The standardized partial regression coefficient of .95 shows that the influence of Category Structure on verification performance is strong indeed! This rather startling degree of predictability from a collection of paper-and-pencil measures completed in a classroom to choice RT performance obtained under highly standardized laboratory conditions alerts us to the potential power of the SEM approach. Compare this regression coefficient of .95 to the zero order Pearson product-moment correlations of .56, -.49, .69, and .49 between, respectively, Association Frequency, Semantic Distance, Typicality, and Featural Overlap, and RT in these same data.

There are some other noteworthy points concerning the estimation procedure. The sets of directed relations on both sides of Figure 1 are simultaneously derived so as to maximize the predictability of Category Verification from Category Structure. Thus, we can conclude from the fact that Typicality is the principle most strongly related to Category Structure that Typicality also is the best single predictor of Category Verification. From the rules of path analysis, we know that the magnitude of this prediction is  $.82 \times .95 = .78$ . Still employing the rules of path analysis, we can account for all of the standardized variance in, say RT, by computing  $(.85 \times .85) + .28 = 1.00$ .

Our next task is to evaluate the adequacy of the model as a whole. As specified in Figure 1, the model does fit the data rather well. In addition, the first-order derivatives (supplied by LISREL-IV) for each of the potential parameters of the model are uniformly low, thus indicating no local areas of lack of fit in the model.

In many cases, the chi square statistic may lead to rejection of the model. The LISREL-IV program provides information on the loci of lack-of-fit that permits one to change the model to improve fit. Ideally, the investigator would next collect fresh data and attempt to confirm the revised model. In actual practice, the model is often "fixed" to improve fit based on preliminary attempts to fit the model. In fact, the three undirected relations in Figure 1 were added to the model in this fashion.

Within the constraints of our data, we can pit rival structural equation models against one another. The result is a test of the relative ability of the two alternate theories to account for the observed covariation. One model we were interested in specified that the four categorization principles should contribute equally to the Category Structure latent variable, i.e., that the four partial regression coefficients be constrained to be equal during the maximum likelihood estimation process. When we evaluated this model, we found a chi square of 51.88 with 10 df. In an opposing model, all conditions were equivalent, except that the four principles of categorization were allowed to vary freely in their estimated contribution to the latent variable. This latter model yielded a chi square of 42.56 with 8 df. Neither model fits very well, but achieving optimal fit is not the purpose of this comparison. A statistical comparison

of the two models shows that  $\chi^2_{diff} = 9.32$  with 2 df, a significant difference ( $p = .01$ ). This comprehensive test of models offers evidence that the four principles are not equivalent measures of Category Structure.

### Concluding Remarks

One goal of this paper has been to convey something of the potential of the SEM approach for model building and testing within the realm of cognitive psychology research. In attempting to do this, we have underemphasized the difficulties of the approach. Perhaps it is appropriate to conclude with some cautions, that are discussed more fully by Horn and McArdle (1980). Given that SEM, like any multivariate maximization procedure, capitalizes on chance relationships in the data, replication of complex findings is mandatory. Issues of identification can be quite intractable, so much so that studies not initially conceptualized with SEM in mind are often unsuitable for SEM analysis. On the other hand, SEM has potentials we have not explored; for example, interactive terms can be entered into the systems of equations and multiple groups of subjects or items can be analyzed simultaneously. On balance, the approach is worth the attention of a discipline that utilizes complex models and is in need of methods for testing them comprehensively.

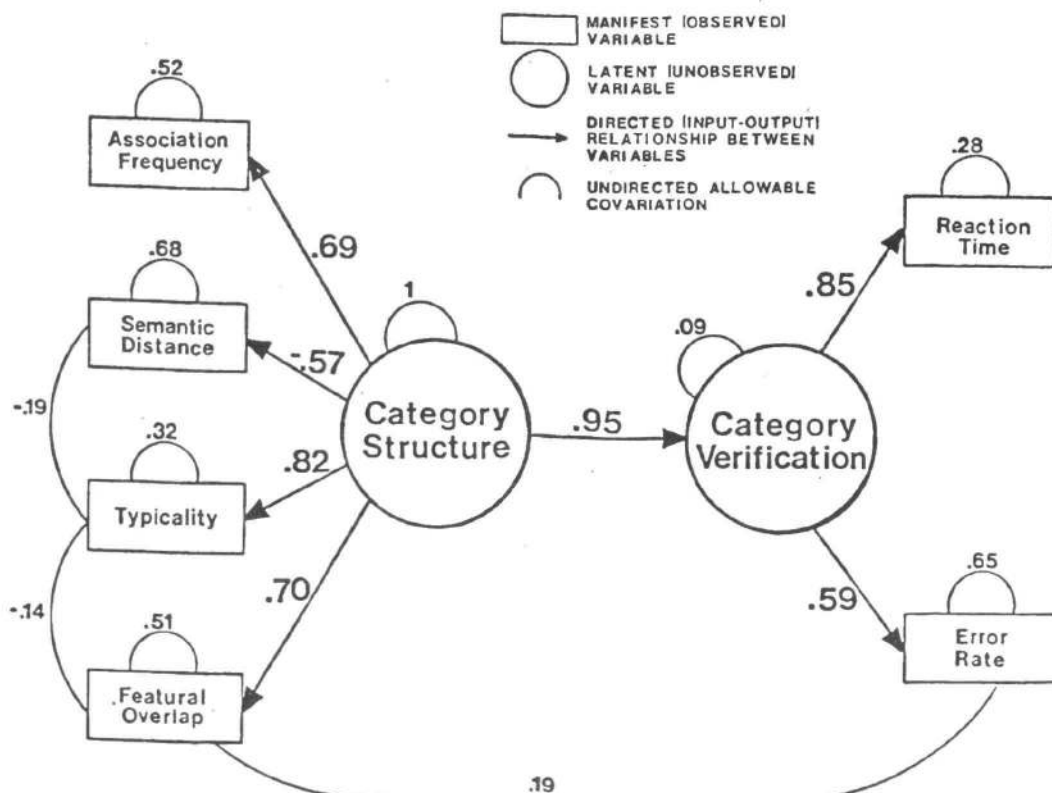
Gernsbacher, M. A. On the use of a "new" performance variable to measure cognitive processing. Manuscript submitted for publication, 1982.

Horn, J. L., & McArdle, J. J. Perspectives on mathematical / statistical model building (MASMOB) in research on aging. In L. W. Poon (Ed.), Aging in the 1980's: Psychological issues. Washington, D. C.: American Psychological Association, 1980.

Joreskog, K. G., & Sorbom, D. LISREL IV: Analysis of linear structural relationships by the method of maximum likelihood. Chicago: National Educational Resources, 1978.

McArdle, J. J., & McDonald, R. P. A simple algebraic representation for structural equation models. Unpublished manuscript, University of Denver, 1981.

FIGURE 1. A Structural Equation Model of the Relationship Between Category Structure and Category Verification



Goodness-of-fit:  $\chi^2 = 3.89$ ,  $df = 5$ ,  $p = .57$ . Standardized estimates are shown. All values are at least twice the size of their standard errors.



COUNTPLAN: A Model for Planning Counting Procedures based on  
Utilization Knowledge, and Procedural and Conceptual Competence

Donald A. Smith and James G. Greeno  
University of Pittsburgh

We are attempting to reach a clearer understanding of the distinction between competence and performance. By "competence" we mean the general knowledge that enables specific occurrences of successful performance. Competence includes understanding of general concepts and principles of a task domain; we refer to this as conceptual competence. For example, performance of young children in counting tasks provides evidence that they understand principles of number such as cardinality and one-to-one correspondence (Gelman & Gallistel, 1978). Judgments about a person's conceptual competence are problematic; the person may understand a principle adequately for a task, but lack skills or situational resources needed to apply the principle.

We report an analysis of knowledge needed to apply conceptual principles of number to perform correctly in counting tasks. Knowledge for application includes understanding of significant general principles, which we call procedural competence, as well as knowledge that applies to features of the specific task setting, which we call utilization knowledge. Procedural competence is understanding of general principles involving relations among goals, actions, and conditions for performance. Utilization knowledge enables features of the task setting to be used in satisfying required conditions and goals.

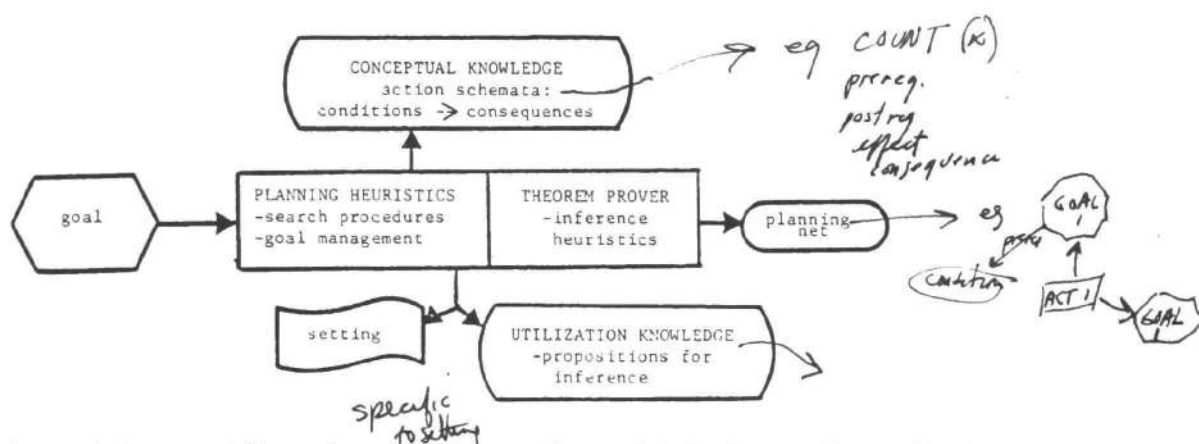
Our analysis uses a formulation of conceptual competence for counting given by Greeno, Riley, and Gelman (in press). Principles of number are represented as a set of schemata that specify requisite conditions and consequences of actions at several levels of generality (cf. Sacerdoti, 1977). The relation of the principles to performance in different situations is represented by planning nets that provide structural descriptions of procedures for counting (cf. VanLehn & Brown, 1980). The



principles, expressed as action schemata, are used as premises of the derivations in a way that is analogous to the rewrite rules used to derive sentences in generative grammars.

In the present analysis, procedural competence and utilization knowledge are cognitive components needed to construct the derivations of planning nets for counting procedures. Figure 1 shows the main components of a production system, COUNTPLAN, that we have implemented. There is a planner that receives goals and constructs planning nets, using two sources of knowledge. One source contains the system's conceptual competence: principles of number in the form of action schemata. The other source is a representation of the task setting, coupled with utilization knowledge that enables inferences that link features of the setting with conditions that are required according to the action schemata. Procedural competence is represented by the heuristics of planning, including a standard means-ends method, procedures for managing goals during top-down planning, and knowledge for co-ordinating goals and actions relating to sets and individual objects. Procedural competence also includes heuristics that can prove theorems about goals and conditions based on features of the setting and propositions included in the utilization knowledge.

FIGURE 1: Components of COUNTPLAN



Planning begins with specification of the goal to find the number of objects in a set, and proceeds by top-down refinement. The planner searches among the action schemata in its conceptual knowledge for one whose consequence matches the current

goal. When one is found, the schema is instantiated and its requisite conditions are formulated as planning goals that can be satisfied either by utilization of the setting or by other actions. Goals set by the planner are of three kinds: goals to enable actions, goals to achieve states corresponding to requisite conditions, and logic goals that arise from quantifiers (e.g., FOR-ALL) and from connectives (e.g., IFF). Planning continues until all goals can be satisfied by a verified plan. Principles of number are reflected in the requisite conditions of the schemata in the conceptual knowledge base, so that satisfaction of those conditions ensures that derived procedures conform to the principles.

The propositions in COUNTPLAN's utilization knowledge are inference rules that transform specific information in the setting into a form that is compatible with the requisite conditions of the goals. In one of the settings that COUNTPLAN encounters the objects that comprise the set of things to count (TTC) are arranged in a straight line. This feature of the setting enables the inference that TTC is ordered. The fact that TTC is ordered allows the further inference that there is a first object and that each subsequent object is connected via a next relationship. Once generated, these properties of an ordered set allow the planner to verify the prerequisites of actions that operate on objects in ordered sets. Actions whose prerequisites cannot be verified in a given setting can be removed from the list of available actions. The process of applying these inference rules is a form of theorem proving in which the planning goals are linked with the inferences made about the problem setting. The requirement of linking goals about sets with actions performed on individual objects provided a problem in formulating planning knowledge that led to interesting insight about procedural competence. In general, the system establishes a global goal to count a set of objects regardless of their exact descriptions. This abstract representation of the object is propagated down the planning net. At a lower level this abstract representation is tagged with certain properties that provide a description of the specific object. This specification is passed to the lowest level



of the plan where it is compared with the object that was retrieved. If the two specifications match, the retrieval process is accepted.

Procedural competence is represented in COUNTPLAN as a set of planning heuristics that handle the goals that cannot be satisfied through utilization knowledge. These heuristics represent general knowledge about relationships among actions, goals and constraints that enables the generation of a plan that is consistent with the principles governing the task. The planning heuristics provide procedures used in searching for the appropriate actions in conceptual knowledge, determining if goals are achieved, and making decisions about goal management. The model utilizes several types of planning rules to accomplish these tasks. One type of rule records significant changes in the state of the world and notes possible goal conflicts. For example, if two actions are included in a procedure and one action requires that a particular set be empty while the other action requires that the set is not empty, a goal conflict is noted. The information about this conflict can then be used to order the two actions appropriately. Another type of planning rule propagates constraints based on logical and requisite relations through the plan. For example, one of the logic goals (FOR-ALL) requires that a plan must assure that all the objects are counted. This constraint is propagated through the plan by linking each object related state and action to the FOR-ALL logical goal structure. These links prevent the system from accepting a plan that does not satisfy the global constraint imposed by the FOR-ALL goal. The links also provide a means of accessing these plan components if they cannot be adequately modified. An additional set of planning rules is responsible for verifying that state goals are true and testing that all the requisite conditions of actions goals have been verified. The final two types of planning heuristics involve monitoring the effects of actions and the constraints imposed by corequisites. One benefit of effects monitoring is that it allows the system to notice that the effect of one action corresponds to the prerequisite of the other action. Provided with this information, the system can adopt a least-commitment

strategy and temporarily suspend the second action rather than making a stronger commitment to retract that action. One of the primary functions of corequisite monitoring is that it constrains the system to generate plans that are consistent with the counting principle of one-to-one correspondence between objects counted and numbers used. The constraints associated with this principle assume a special status within the system and the detection of a violation shifts the attention of all subsequent planning to restoration of a state of balance.

The generative capacity of the theory has been examined by analyzing several counting tasks that impose different constraints on the plan. These analyses demonstrate that the model is flexible in the sense that procedures are planned for various arrangements of objects, and robust in the sense that existing procedures can be modified to satisfy additional constraints. The analyses contribute toward development of a theory of implicit understanding by providing a mechanism to analyze the relationships between formal principles relevant to a task and cognitive procedures for performance in the task.

utilization  
proceeds  
procedural  
util < proceed < comp  
References

robust

Gelman, R. & Gallistel, C.R. The child's understanding of number. Cambridge, Mass.: Harvard University Press, 1978.

Greeno, J.G., Riley, M.S. & Gelman, R. Young children's counting and understanding of principles, Cognitive Psychology, in press.

Sacerdoti, E.D. A structure for plans and behavior. North-Holland, Amsterdam: Elsevier, 1977.

VanLehn, K. & Brown, J.S. Planning nets: A representation for formalizing analogies and semantic models of procedural skills. In R.E. Snow, P.A. Frederico, & W.E. Montague (Eds.), Aptitude, learning, and instruction (Vol. 2): Cognitive process analysis of learning and problem solving. Hillsdale, N.J.: Lawrence Erlbaum, 1980.

Driving a different car  
eg making a different  
planning net.

Generative:  
small set of ppls, in  
plans of procedures

Flexibility: same  
ppls in different settings  
eg driving a different  
car (this is robustness)

### FUNDAMENTAL PPLS

STABLE ORDER

CARDINALITY: last elt in name  
set size

1-1 CORRESPONDANCE

INDIFFERENCE OF ORDER

INDEPENDENCE FROM THINGS  
COUNTED

CONCEPTUAL  
PROCEDURAL  
UTILIZATION

COMPETANCE: PPLS  
ACTION  
SCHEMA

robustness  
modify  
procedures to fit  
new situations



# **Paper Session #4**



## COMPREHENDING WORD ARITHMETIC PROBLEMS:

### A PSYCHOLOGICAL PROCESSING MODEL

Walter Kintsch

University of Colorado

A general theory of discourse comprehension (van Dijk & Kintsch, 1983) is used to develop a model for how people understand and solve word arithmetic problems, incorporating a problem solving model proposed by Riley, Greeno, & Heller (1982). This work is being done in collaboration with James Greeno, University of Pittsburgh.

Understanding mathematical word problems is a very special kind of understanding. Specialized strategies are involved (which have to be taught specifically in school), and the textbase that is constructed is a peculiar one, but its peculiarity lies in its specialized content - the same kind of structures are being generated as in any other situation, people reading a newspaper, a story, or a textbook. Thus, we are not proposing to build a specialized comprehension front-end for a word-arithmetic problem-solving model; instead, we are going to apply a general model of text comprehension to this special situation. Specifically, what the theory assumes is that the verbal input is decoded into a list of atomic propositions which are organized into larger units on the basis of some knowledge structure to form a coherent textbase. From this textbase, a macrostructure is constructed which represents the most essential information in the textbase. In parallel with this hierarchical text representation we also construct a situational model, which in this case is the problem representation which Riley, Greeno & Heller (1982) used as the starting point for their model, and upon which various arithmetic operations can be performed.

A set of problems can be constructed that form prototypes for all single-step addition and subtraction problems. By suitably restricting the language of these problems, only 9 propositional frames must be used in these problems, which makes the task of deriving atomic propositions from the problem sentences easy, indeed trivial. Each proposition is associated with a meaning postulate, and it is at this point where arithmetic-specific aspects enter into our analysis, because the meaning postulates used here are quite special ones. They are impoverished compared with everyday language use, and they are specialized. All we care about in these problems concerns sets of objects (always marbles here), their specification (always in terms of ownership), their numbers, and the relationships among the various sets. Thus, the only kind of information that is relevant here is that specified by the slots of the set schema. The textbase is always formed from this set schema. This is very different from other types of texts, where the textbase may be based on many diverse knowledge structures, requiring a richer interpretation. If we read "Joe gave 5 marbles to Tom" in an arithmetic problem, all we want to know is that there is a set of marbles now owned by Tom and formerly owned by Joe, which we call a transfer set and which is part of a Transfer schema, together with a startset and a resultset. In a story, on the other hand, we might be concerned with Joe's motive, or with Tom's reaction, or we would prefer dollars to marbles - all of which would be out of place here.

Thus, sentences are decoded into atomic propositions and these are assigned to the slots of a set schema. The main purpose of the model is to show exactly how this happens. The basic assumption is that the process is strategic and that the strategies involved are not the "normal" comprehension strategies, but specialized strategies for dealing with word arithmetic problems. Formally, strategies are

modelled as productions. That is, we specify certain conditions in the text which, if they occur, lead to certain actions. The actions always consist in constructing a set and assigning the text propositions to its slot. We need five such schematic strategies or productions to account for the problems considered here. In each case, we find on the condition side a quantifier proposition of the form  $N(y)$  where  $N$  is a number or SOME, and  $y$  are some marbles, plus either a HAVE, GIVE, or MORE/LESS-THAN proposition which provides information on the role of the set to be created. For example,  $N(y)$  in the context of a GIVE proposition creates a transfer set of  $N$  marbles, owned by the patient of the GIVE proposition. MORE (or LESS) propositions are the condition for creating a remainder set. Simple HAVE propositions, on the other hand, provide no information about the role of the set. In this case, the role slot remains empty, until there is other contextual information that permits filling it. In the Change problems, sentence order serves that function, as well as such explicit propositions as NOW, THEN, or PAST which further specify temporal order and hence the role of sets in the transfer schema. In other problems, there may be no linguistic indicators of set role at all, and hence, the role slot is not filled in the text base. However, the role slot must always be inferred in the problem representation which is constructed in parallel with the text base. Thus, the problem representation is in part a copy of the text base, except for those schema slots which the text base does not specify and which must be inferred from the knowledge of some higher order schema. For instance, consider the combine problem: Joe has 3 marbles. Tom has 5 marbles. How many marbles do they have altogether? The first two sentences provide the conditions for creating two sets of marbles, owned by Joe and Tom, with a certain number in each, but with unknown roles. The third sentence triggers a MAKE-SUPERSET because of the HOWMANY(MARBLES) in the context of HAVE(JOE & TOM, MARBLES). Having a superset, we need subsets, and we infer that the two sets created previously are the subsets in question: their specification in terms of ownership permits us to make this inference, but it does not force us to do so logically - the inference is only justified within the pragmatic conditions of word problems. The inferred roles are specified in the problem representation, but no corresponding proposition "S1 is a subset" is inferred in the textbase. Thus, the problem representation may contain more information than was explicit in the textbase. The reverse may also be true: if we had included in our problem the irrelevant information that Tom had blue marbles, the proposition BLUE(MARBLES) would have been assigned to the specification slot of the corresponding set in the textbase, but would not affect the problem representation.

Once the problem representation is completed, the arithmetic operations themselves are performed, the different set constellations serving as the conditions for appropriate operations. Thus, for instance, a transfer-in schema with a calculation goal on the result set is the condition for a count-on operation in young children, or addition in the older.

How all this works to produce the right solution to a word problem is best illustrated by a few examples. However, because of space limitations, we can only describe here a particularly simple example, the Change 1 problem: Joe has 3 marbles. Tom gives him 5 more marbles. How many marbles does Joe have now?

The first sentence is "Joe has three marbles". It is parsed into the propositions P1-P4 as described in Kintsch (1982). P3 and P4 turn out to be the condition for a MAKE-SET operation, creating S1: the four propositions are assigned to the appropriate slots of the set schema. At the same time, a parallel set is established in the problem representation, with entries derived from the textbase (indicated by arrows). Note that at this point there are no entries in the Role slot of S1, neither at the textbase level nor at the level of the problem representation.



The second sentence is similarly organized into S2 via a MAKE-TRANSFERSET operation, while S1 is held in short-term memory. Since S2 is a transferset, requests are created for the corresponding start- and resultsets. S1 is identified as the desired startset on the basis of an explicit linguistic cue, the "then" of the second sentence. Short-term memory now contains a partially completed transfer schema consisting of S1 and S2, and a request for the missing resultset.

The third sentence provides this resultset and completes the schema, which then triggers the arithmetic operation count-on (or add), as in the Riley model.

All the other problems can be treated similarly, using the propositional schemata for the construction of the proposition lists, and the schematic strategies to organize them into TRANSFER, SUPERSET, or MORE-THAN schemata. However, the process does not always run off as smoothly as for Change 1 problems: sometimes, inferences need to be made to specify a slot of a schema for which the text provides no explicit cues (an example was mentioned above for Combine problems) and sometimes sets no longer available in the limited-capacity short-term memory buffer must be reinstated from long-term memory (or by rereading a sentence) to complete a problem.

Thus, our model leads us to distinguish three separate sources of problem difficulty. In order to do these problems, you need first of all knowledge about the right arithmetic schemata and operations - the set schema, the transfer, superset, and more-than schemata, as well as the actual counting and arithmetic operations; in this respect, our model is no different than the Riley et al. model in its implications. But you also need to be able to use these knowledge schemata in word problems, i.e., you need to have this knowledge in the form of productions adapted to the textual input. Finally, even if you have all the right knowledge, the way a problem is stated may make it easy or hard because some problem versions make only minimal demands on short-term memory (e.g., Change 1 or Combine 1) while others can only be solved if large amounts of sometimes incoherent material can be remembered.

Preliminary analyses have shown that the need to make inferences, and especially the short-term memory load (the size of the units to be carried in STM, the number of requests that must be kept track of, and number of propositions that can not immediately be attached to some set-unit) are factors which greatly contribute to problem difficulty, over and beyond the knowledge structures and schematic productions needed to solve the various problem types.

#### References

- van Dijk, T.A. & Kintsch, W. Strategies of comprehension. New York: Academic Press, 1983.
- Greeno, J.G., Riley, M.S., & 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, 1983.
- Kintsch, W. Aspects of text comprehension. In J.F. LeNy & W. Kintsch (Eds) Language and comprehension. Amsterdam: North-Holland, 1982.



# Children's Mental Models of Recursive Logo Programs

by

D. Midian Kurland & Roy D. Pea  
Center for Children and Technology  
Bank Street College of Education

Abstract. Children with a year of Logo programming were asked to think-aloud about the function of some Logo recursive programs, and then to predict by hand-simulation of the programs what the graphics turtle will draw when the program is executed. If discrepancies arose, children were asked to explain them. A prevalent but misguided "looping" mental model of Logo recursion persisted even in the face of contradictions between program effects and the child's predictions.

Introduction. The power and beauty of recursion as a development in the history of programming languages (such as LISP and Logo) and its conceptual importance in mathematics, music, art and cognition generally is widely acknowledged (4). Less attention has been given to the developmental problem of how people learn to use the powers of recursive thought and recursive programming procedures. Our approach to this question is influenced by several findings basic to a developmental cognitive science, specifically, the role of mental models in guiding learning and problem solving, and the widespread use of systematic, rule-guided problem solving approaches by children, not only adults (10). Understanding recursive functions in programming involves notational and conceptual problems, the latter including problems with understanding flow of control and data. Expert programmers are guided by a valid mental model of how program code controls computer operations. Novices' faulty models are adapted in response to direct instruction and feedback from their own programming and debugging experiences, in which conflicts between their current model and program behavior is reflected upon.

A widespread belief among computer educators is that young children can "discover" the powerful ideas formally present in programming simply through experimenting within a rich programming environment, as if unconstrained by prior understandings. This belief is largely due to Papert's (7) popular account of Logo, a LISP-like language designed for children to allow them to develop powerful ideas, such as recursion, in "mind sized bites". Many assume children can learn recursion through self-guided explorations of programming concepts in Logo. However, our observations of 8-12 year-olds indicate that most avoid all but simple iterative programs, which do not require the deep understanding of control structure prerequisite for an understanding of recursion.

In a study examining children's ability to develop recursive problem descriptions, Anzai & Uesato (1) have shown how adolescents' understandings of recursive formulations of the factorial function is facilitated by a prior understanding of iteration. They demonstrate that for mathematics, recursion can be learned via a discovery process by most children, particularly if they have first experimented with iterative functions. Of their subjects who correctly identified iterative structure in a set of problems, 64% were also able to work out recursive solutions to a second problem set. However, only 33% of subjects who did not have prior iteration experience worked out the recursive functions. Anzai & Uesato conclude that understanding recursion is aided by an understanding of iteration, but urge caution when extending this point "to more complex domains such as computer programming ... [since] a complex task necessarily involves many different cognitive subprocesses, and it is not always easy to extract from them only the part played by recursion" (p. 102). While Anzai & Uesato focus on the insight necessary to generate a recursive description of a math function, in programming one must acquire that insight and be able to implement it in specific programming formalisms. In addition to understanding recursion, the child must understand the logic and terminology governing the language's control structure. Adult novices have trouble with both. Learning to program they have great difficulties in thinking through flow of control concepts such as Pascal's while loop construction (9) and tail recursion in SOLO, a Logo-like language (5), even after extensive instruction. Furthermore, Bonar (2) finds that prior natural language understandings

of programming terms misleads novice programmers in their attempts at explaining how a program works. Prior meaning is brought to the task of constructing meaning from lines of programming code. We expect children will also be guided in their interpretation of programming language constructs by their natural language meanings, and by faulty mental models of flow of control structure. Indeed, a common lament of programming instructors is that novices have great trouble acquiring the concept of recursion and the ability to use recursive formalisms in their programs.

#### How recursion works in Logo: A user's perspective

When a Logo program is run, if a procedure references itself, execution of that procedure is temporarily suspended, and control is passed to a copy of the named procedure. Passing of control is active when the programmer explicitly directs the program to execute a specific procedure. However, when the execution of this version of the procedure is finished, control is automatically passed back to the suspended procedure, and execution resumes at the point where it left off. Passing of control is passive here because the programmer did not need to specify where control should be passed in the program.

To understand how recursive procedures work in Logo one must know:

(1) The rule that execution in Logo programs proceeds line by line. However, when a procedure calls another procedure or itself, this inserts all lines of the named procedure into the executing program at the point where the call occurred. Control then proceeds through each of these new lines before carrying on with the remaining lines of the program. Thus control is passed forward to the called procedure, and then is passed back to the calling procedure.

(2) That when a procedure is executed, if there are no further calls to other procedures or to itself, execution proceeds line by line to the end of the procedure. The last command of all procedures is the END command. END signifies that execution of the current procedure has been completed and that control is now passed back to the procedure from which the current one was called. END thus (1) signals the completion of the execution of one logical program unit, and (2) directs flow of control back to the calling procedure so the program carries on.

(3) That there are exceptions to the line by line execution rule. An important one for recursion is the STOP command. STOP causes the execution of the current procedure to be halted, and control to be passed back to the procedure from which the currently executing one was called. Functionally, then, STOP means to branch immediately to the nearest END statement.

How well novice programmers' mental models of the workings of recursive procedures took into account these three central points was our research focus.

Subjects. Seven children (2 girls and 5 boys, 11-12 years old) in their second year of Logo programming participated in the study. The children were highly motivated to learn Logo programming, and had averaged over 50 hours of classroom programming time under the supervision of experienced classroom teachers knowledgeable in the Logo, and who by choice followed Papert's "discovery" Logo pedagogy (7). All children had received instruction in iteration and recursion, and had demonstrated in their classroom programming that they could use iteration and recursion in some contexts.

Materials. Short Logo programs were constructed of procedures which reflected four levels of complexity: (1) procedures using only direct commands to move the turtle; (2) procedures using the iterative REPEAT command; (3) tail recursive procedures; and (4) embedded recursion procedures. This paper focuses on the revealing features of children's performance at levels 3 and 4. Examples of programs at levels 3 and 4 are (:SIDE = 80 for each):

#### Level 3: tail recursion program

```
TO SHAPEB :SIDE
  IF :SIDE = 20 STOP
  REPEAT 4 [FORWARD :SIDE RIGHT 90]
  RIGHT 90 FORWARD :SIDE LEFT 90
  SHAPEB :SIDE/2
END
```

#### Level 4: embedded recursion program

```
TO SHAPEC :SIDE
  IF :SIDE = 10 STOP
  SHAPEC :SIDE/2
  REPEAT 4 [FORWARD :SIDE RIGHT 90]
  RIGHT 90 FORWARD :SIDE LEFT 90
END
```



### Experimental procedure

Our choice of a method was guided by comprehension studies which utilize "runnable mental models" (3) or simulations of operations of world beliefs in response to specific problem inputs. Children were asked to think aloud about how a Logo procedure would work, then to hand simulate the running of each program line by using a turtle "pen" on paper. Then they were shown the consequences of running the program they had explained, and if their simulation mismatched the turtle's actions, they were asked to explain the discrepancies, and one additional problem at that level was presented.

### Results

All seven children made accurate predictions for programs at the first two complexity levels with only minor difficulties. They expressed no problems with the recursive call of the tail recursive programs of level 3; however, two children treated the IF statement as an action command to the turtle, and another assumed that since she did not understand the IF statement the computer would ignore it. No child made accurate predictions for either embedded recursion program at level 4. The children's problems with explaining embedded recursion may be traced to two related sources. The first involves general bugs in their mental model for how lines of programming code dictate the computer's operations when the program is executed, while the second concerns the particular control structure of embedded recursive procedures.

#### (1) General bugs in program interpretation

Decontextualized interpretation of commands. Children carried out "surface readings" of programs during their simulations. They attempted to understand each line of programming code individually, ignoring the context provided by previous program lines. They stated each command's definition rather than treating program lines as parts of a functional structure in which the purpose of particular lines is context-sensitive and sequence-dependent. This led to trouble during their simulations in keeping track of the current value of the variable SIDE, and in determining the actual order in which lines of code would be executed. Understanding recursion is impossible without this knowledge about sequential execution. The child must learn to ask: "How does the line I'm reading relate to what has already happened and affect the lines to follow?" The two bugs which follow concern an opposite tendency, an overrich search for meaning in other program lines.

Assignment of intentionality to program code. Children often did not distinguish the meaning of a command line they were simulating from the meaning of command lines they expected to follow (e.g. lines that if executed would draw a BOX). For example, in program SHAPEC, one child said of the IF statement: "If :SIDE equals 100 stop. O.K., I think this will make a box that has a hundred side." Another child at the same point said: "this makes it draw a square."

Treating programs as conversation-like. As in understanding conversation, and in problems the non-schooled encounter in formal reasoning (where beliefs about the truth of an argument's premises are focused on rather than the validity of its form: (6), (8)), children appropriate for problem solving any knowledge they believe will help them understand. In the case of Logo program comprehension, this empirical strategy has the consequence of "going beyond the information given" to comprehend the meaning of lines of code, such as deriving implications from one code line (e.g. an IF statement) about the meaning of another line. For example, one child interpreted the recursive statement in SHAPEC as having the intention of drawing a square, predicting that the turtle would immediately draw a square before proceeding to the next command.

Overgeneralization of natural language semantics. Children interpreted the Logo commands END and STOP by natural language analogy, leading them to believe that when the terms appear the program completely halts. Several children concluded that SHAPEC would not draw at all, since when :SIDE reaches the value of 10, the program "stops, it doesn't draw anything." In fact, STOP and END each passively return control back to the most recently active procedure, and drawing occurs.

Overextension of mathematical operators. Children expressed confusion about the functions of numbers as inputs, and in arithmetic functions such as dividing the variable value, or addition of a constant to it, during successive procedure calls.

For example, one child explained SHAPEC this way:

...if SIDE equals 10 then stop. See, instead of going all forward 80, you just go forward 10. Then you're gonna stop. Then you're gonna go. Then (line 3) I guess what you're gonna do is keep on repeating that 2 times, so it'd be forward about 20 instead of forward 10, forward 20 (line 4), and you're gonna repeat 4, so it'd be forward 80 because it says repeat 4 forward side...

Numbers were also often pointed to as the mysterious source of discrepancies between the child's predictions and the results of program execution.

(2) Mental model of embedded recursion as looping.

The children were fundamentally misled by thinking of recursion as looping. While this mental model is adequate for active tail recursion, it will not do for embedded recursion, which requires an understanding of both active and passive flow of control. The most pervasive problem for all children was this tendency to view all forms of recursion as iteration. For example, one child explained the recursive call in program SHAPEB in the following manner:

[the child explained what the first four lines did, then said]: "line 5 tells it to go back up to SHAPE, tells it to go back up and do the process called SHAPEB, this is the process [points to lines 2-4]. It loops back up, and it divides SIDE by 2 so then SIDE becomes 40...[carries on explaining correctly that the procedure will draw two squares]"

In this example, the child clearly views tail recursion as a form of looping, rather than as a command to suspend execution of the currently executing procedure and pass control over to a new version of SHAPEB. However, in this case his wrong model leads to the right prediction, so he is not compelled to probe deeper into what the procedure is doing. This same child explained that SHAPEC:

"...checks to see if SIDE 80 equals 10. If it does, end the program. Next, line 3 [the recursive call] tells it to go back to the beginning except to divide SIDE by 2 which ends up with 40. Then it goes down there (line 2) checks to see if SIDE is 10...[then] back to the beginning...[continues to loop back until SIDE equals 10 then] checks to see if it equals 10, it does, stops. OK, a little extra writing there (points to lines 4 and 5). [draws a dot in the paper to indicate his prediction of what the procedure will do and comments] and that is about as far as it goes because it never gets past this SHAPE (line 3). It is in a loop which means it cannot get past 'cause every time it gets down there (line 3), it loops back up."

This time the child's explanation and prediction were incorrect since SHAPEC makes the turtle draw a series of three squares in a line, each twice as big as the previous one. The child expressed complete bewilderment when the procedure was executed, and could offer no explanation to account for the discrepancies. On the second program of this type, which draws three squares of different sizes inside one another, the child worked down to the recursive call and then said:

"um. wait a minute. I don't understand this. Well anyway, from past experience, like just now, I guess it's not going to listen to that command (points to the recursive call) and it's going to go past it, and it's going to [draw a square] and I guess its going to end then."

Again, when the procedure was run and the child saw he was wrong he expressed confusion, but instead of looking for an error of understanding, he asked:

"Is this the same language we used last year? Because last year if you said SHAPE, if you named the program in the middle of the program, it would go to that program. We did that plenty of times, but it's not doing that here. I don't know why."

The child blamed the language for not conforming to his expectations, but in doing so he indicated that at some level he knew the correct meaning of a recursive call: "it would go to that program." However, though he seemed to know the rule, when he worked through a program, his simpler, and in many cases successful, looping model prevailed.

Discussion and conclusions. We believe these findings are important because they reveal that the children's conceptual bugs in thinking about the functioning of recursive computer programs are systematic in nature, and the result of weaker

theories that do not correspond to procedural computation in Logo.

These findings also imply that, just as in the case of previous work with adults, programming constructs often do not allow mapping between meanings of natural language terms and programming language uses of those terms. Neither STOP or END stop or end, but pass control back. This is important for the Logo novice because when their mental model of recursion as looping fails, they have no way of inferring from the syntax of recursion in Logo how flow of control does work. So they keep their inadequate looping theory, based on their successful experience with it for tail recursion, or blame discrepancies between their predictions and the program's outcomes on mysterious entities such as numbers, or the "demon" inside the language itself. An important issue of a development theory of programming then is: How do inadequate mental models get transformed to better ones?

For a developmental psychology of programming, we require an account of the various factors that contribute to learning central computational concepts. So far efforts to help novices learn programming languages through utilizing programming tutors or assistants have bypassed what we consider to be some of the key factors contributing to novice's difficulties working with computational formalisms. Beyond mistaken mental models about recursion, we have found these to involve atomistic thinking about how programs work, assigning intentionality and negotiability of meaning as in the case of human conversations to lines of programming code, and application of natural language semantics to programming commands. In studies underway, it appears that none of these sources of confusion will be intractable to instruction, although their pervasiveness in the absence of instruction, contrary to Papert's idealistic individual "Piagetian learning", suggests that self-guided discovery needs to be mediated within an instructional context.

#### Acknowledgements

This work was supported by the Spencer Foundation. We wish to thank participants of a workshop at MIT's Division for Studies and Research in Education, from Geneva and Cambridge, for provocative discussions of these issues. Sally MacKain provided invaluable assistance in running the study and providing transcripts.

#### References

- (1) Anzai, Y. & Uesato, Y. Learning recursive procedures by middle school children. Proceedings of the Fourth Annual Conference of the Cognitive Science Society. Ann Arbor, August 1982.
- (2) Bonar, J. Natural problem solving strategies and programming language constructs. Proceedings of the Fourth Annual Conference of the Cognitive Science Society. Ann Arbor, August 1982.
- (3) Collins, A. & Gentner, D. Constructing runnable mental models. Proceedings of the Fourth Annual Conference of the Cognitive Science Society. Ann Arbor, August 1982.
- (4) Hofstadter, D. R. Gödel, Escher & Bach: An eternal golden braid. New York: Vintage Books, 1979.
- (5) Kahney, H. & Eisenstadt, M. Programmers' mental models of their programming tasks: The interaction of real-world knowledge and programming knowledge. Proceedings of the Fourth Annual Conference of the Cognitive Science Society. Ann Arbor, August 1982.
- (6) Luria, A. R. Cognitive development. Cambridge, Mass.: Harvard University Press, 1976.
- (7) Papert, S. Mindstorms. New York: Basic Books, 1980.
- (8) Scribner, S. Modes of thinking and ways of speaking: Culture and logic reconsidered. In Johnson-Laird, P.N. & Wason, P.C. (Eds.), Thinking. Cambridge: Cambridge University Press, 1977.
- (9) Soloway, E., Bonar, J. & Ehrlich, K. Cognitive strategies and looping constructs: An empirical study. Comm. ACM, 1983, in press.
- (10) Siegler, R. S. Developmental sequences within and between concepts. Monog. of the Society for Research in Child Development, 1981, 46 (Serial No. 189).





# HOW DO CHILDREN THINK ABOUT NUMBERS?

Let us count the ways

James A. Levin

Marcia J. Boruta

University of California, San Diego

Andrea L. Petitto

University of Rochester

## Abstract

Children use a wide variety of strategies when dealing with numbers. This diversity has previously been approached through a study of "bugs", strategies that are flawed in various ways. However, we have uncovered a variety of valid strategies which different children bring to bear when dealing with numbers in simple addition and subtraction tasks. By observing these strategies in computer-based estimation games, we have identified some of the components of expert mathematical knowledge.

## Introduction

What makes a person an expert mathematical problem solver? The uniformity of the current mathematics curriculum and many of the current psychological models of mathematical problem solving imply that there is one way to represent numbers and one set of processes for solving a given mathematical problem.

Our observations of children dealing with a computer math microworld have highlighted instead the diversity of approaches they bring to bear. Some of the approaches are "buggy", such that the results they produce are systematically incorrect. However, others are quite different from the standard approaches children are taught in school, yet are equally valid representations of number and numerical operations.

## Shark Shooting

One computer microworld we have developed and used is called "Moving Shark". In this world, children see the fin of a shark on the video screen, then the fin disappears beneath the water, leaving a set of ripples. The children see a "digital readout" at the top of the screen which tells how far the shark has moved underwater. They are to type in a number saying where to throw a harpoon to hit the hidden shark.

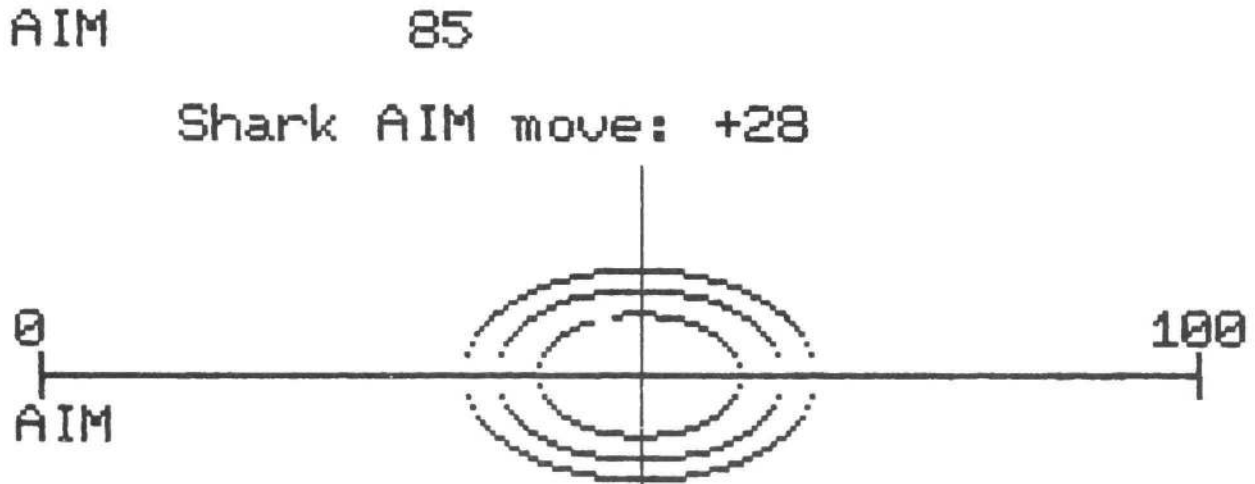


Figure 1: Moving Shark Microworld

This example is an addition task: the player has to add the movement amount to the initial position in order to compute the new position. One common approach, which most adults employ, goes like this: 1) estimate the number specifying the initial position (in this case, perhaps 55), 2) add to that number the movement number, and 3) type in the sum.

This approach is so natural and obvious that we might dismiss all others as "buggy" unless we kept an open mind. Another approach, which we've seen children apply to this problem is: 1) estimate the length of movement represented by the movement number, 2) apply that length on the screen to the initial location, 3) estimate the number specifying the new position (in this case, perhaps 85).

In this little microworld, there are two very different representations of numbers: as decimal digits and as position on a number line. The goals embedded in the game require dealing with both of these ways of thinking about the same thing. In fact, simpler versions of this game mainly require translating from position on a number line to digits (a game we developed called Harpoon) or from digits to position on a number line (a game we developed called Sonar). We have found that children differ on their facility in each of these two ways of thinking about number.

With each way of thinking about numbers, there are corresponding numerical processes for manipulating numbers. So, addition of digits is the multi-step right-to-left symbolic algorithm that we all learned as "addition". However, addition of two numbers can be carried out by representing the two numbers as positions on a number line, then translating the line length representing one number to the end of the other. Galton (1907) reported this technique as the way that one of his subjects standardly did addition and subtraction. Similar estimation techniques have been discovered with subjects doing multiplication (Lopes, 1976).

What we have discovered is a variety of "non-standard" (but mathematically valid) ways of representing numbers and numerical operations. Are these non-standard ways just curiosities, illustrating the wonderful perversity of human nature in bucking twelve or more years of the best efforts of the educational establishment?

We find these variations non-trivial for two important reasons:

- 1) They are consistent with close observations of the ways that adults actually deal with numbers.
- 2) They may hold the key to the nature of mathematical expertise.

#### Mathematical expertise

Lave and her associates have been carrying out careful studies of how "just plain folks" in Southern California deal with numbers in their everyday lives (Lave, Murtaugh, & de la Rocha, 1983). For example, they observed housewives doing their grocery shopping. They found little use of the standard multi-place algorithms learned in school. Instead, they found an extensive use of a diverse set of specialized estimation strategies. In many cases, the people were somewhat embarrassed to have these specialized strategies observed, since they felt they "should" use the standard algorithms.

These findings have been reinforced by the studies of Scribner, observing the arithmetic of warehouse workers (Scribner, 1983). Again, these workers use a diverse set of specialized approaches for performing arithmetic operations in service of their work. Before we dismiss these findings with a cavalier conclusion that people are lazy, let us examine some studies of expert physics problem solvers.

Larkin and her associates (Larkin, McDermott, Simon, & Simon, 1980) have been studying the ways that experts and novices differ in solving physics problems. When a physics teacher demonstrates in a class how to solve a problem, s/he often describes the problem, then immediately writes down a set of equations, which, when applied in the correct order, lead to the desired solution. Novices, taking this as a model, start solving the problems given to them by writing equations. They then get stuck in a quagmire of details.

A close examination of expert solvers solving unfamiliar problems shows that there are several preliminary steps that they take which are not often revealed to novices. Experts initially represent the problem to themselves in a qualitative, global way, often by drawing a diagram that abstracts out the major factors. At this point, they can then classify the problem as being an instance of some general type. These preliminary steps then guide the experts in writing the appropriate equations and using them in an order that leads directly to a solution.

The hallmark of an expert is the ability to think about a problem in their domain of expertise in multiple ways, and to draw upon these multiple points of view in a sequence that leads straightforwardly to a solution. This kind of "orchestration" of multiple representations of the problem elements is just what we find distinguishes experts in our estimation microworld from novices.

When a player misses the shark, s/he gets several kinds of information as feedback. First of all, the harpoon "splashes" into the water, leaving a set of ripples to mark the spot. Secondly, the player gets textual feedback, either "Smaller", "Bigger", or "Right On!". Thirdly, an arrow appears that indicates which way to go in making the next guess.

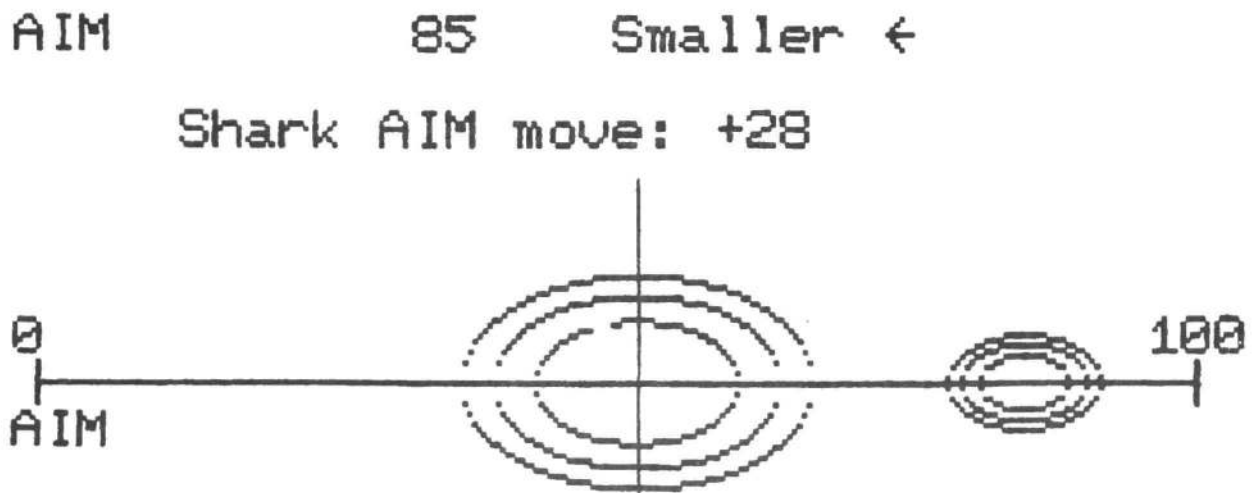


Figure 2: Moving Shark, First Guess

Some of the novice players only pay attention to the textual feedback or to the arrow feedback, effectively turning the game into a bisection strategy game like Huckle. Other players use this same feedback, but instead of bisection, they increment or decrement their guess by a fixed amount (often some multiple of ten). Still other players use the "splash" feedback to calibrate the previous location of the shark, and then reevaluate their addition estimation. Expert players often use several strategies, within games or between games. Within a game, some experts use different strategies for the first guess and for subsequent guesses. Some even use multiple strategies within a single harpoon throw decision.

### Theoretical implications

What are the theoretical implications of this view of expertise? The notion of coordinating multiple simultaneous approaches to a problem is very much in the spirit of the current work on parallel distributed processing (Hinton & Anderson, 1981). But what kind of representation can we have for numbers that allows us to capture very different ways of thinking about numbers and their representation? Most cognitive models have proposed relatively limited representations for numbers, often finessing the entire issue by representing numbers by real numbers. Others have proposed multiple, independent representations for the different aspects. Shepard, Kilpatrick & Cunningham (1975), for example, propose several independent multidimensional scaling spaces to represent the numbers less than 10.

We have proposed a "landmark representation" for continua (Levin, 1981; Hutchins & Levin, 1981). In this representation, there are a set of "landmark" concepts for particular discrete numbers. Any particular number is then represented by the differential "activation" of one or more of these landmarks. This framework allows a substantial diversity of representational types, each with its own set of landmark concepts. These different types are then coordinated by the interaction of the landmarks of the types when simultaneously activated.

The particular types of number manipulation processes (such as addition or subtraction) are cognitive processes that are defined to operate on one or more of these types of number landmarks. So, for example, the multi-place right-to-left digit-based algorithm is defined in terms of the landmarks of the digits; the quantity manipulation algorithm is defined in terms of the spatial landmarks for numbers; etc.

Now we have a way of thinking systematically about the diversity that seems to characterize expert functioning in rich environments. So what? Well, even at this preliminary stage, we can derive from this view some design principles for educational software. If expert functioning depends upon the coordination of many different points of view on a problem, then it may be valuable to present to learners different views of a problem simultaneously, so that they can learn how the different points of view coordinate with each other. This is a feature we have built into many of our educational tools and educational games, and this feature seems to have pedagogical power.

## References

- Galton, F. Inquires into human faculty and its development. New York: E.P. Dutton & Co., 1907.
- Hinton, G. E., & Anderson, J. A. Parallel models of associative memory. Hillsdale, NJ: Erlbaum, 1981.
- Hutchins, E. L., & Levin, J. A. Point of view in problem solving. Proceedings of the Third Annual Conference of The Cognitive Science Society. Berkeley, California, 1981
- Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. Expert and novice performance in solving physics problems. Science, 1980, 208, 1335-1342.
- Lave, J., Murtaugh, M., & de la Rocha, O. Recounting the whole enchilada: The dialectical constitution of arithmetic practice. In B. Rogoff & J. Lave (Eds.), Everyday cognition: Its development in social context. Cambridge: Harvard University Press, 1983 in press.
- Levin, J. A. Continuous processing with multilevel memory representations. La Jolla, CA: Laboratory of Comparative Human Cognition, 1981.
- Lopes, L. L. Model-based decision and inference in stud poker, Journal of Experimental Psychology: General, 1976, 105, 217-239.
- Scribner, S. Studying working intelligence. In B. Rogoff & J. Lave (Eds.), Everyday cognition: Its development in social context. Cambridge: Harvard University Press, 1983 in press.
- Shepard, R. N., Kilpatrick, D. W., & Cunningham, J. P. The internal representation of numbers. Cognitive Psychology, 1975, 7, 82-138.

---

This research was supported by the National Science Foundation, Research in Science Education Grant SED-8112645. Any opinions, findings, conclusions or recommendations expressed are those of the authors and do not necessarily reflect the views of the National Science Foundation. Many thanks to Robert Rowe and Karen Johnson, to the Oceanside Unified School District, and to the Laboratory of Comparative Human Cognition for providing social support.





# A Computational Description of the Stages of Development of Object Identity by Infants\*

by

George F. Luger\*\*, Jennifer G. Wishart\*\*\*, and T. G. R. Bower\*\*\*

## Abstract

This paper is in two parts: I, a brief description of the *identity* theory of the development of the object concept and a computational model that describes the three stages of that theory and, II, the running of that model on two experiments important in the analysis of development of the object concept. The model is designed to help determine which parameters in the development process might best explain changes within the object comprehension skills of infants. The discussion presents a set of perceptual invariants that can be used: a), to describe changes between developmental stages and, b) offer the potential for a cost/gain metric for that development.

The model is written in PROLOG, a very high level computer language. Detailed descriptions of PROLOG (6)(15) and further uses of PROLOG to model problem solving skills (7) may be found elsewhere.

## I. Identity Theory and the Computational Model

The *identity* theory proposed by Bower and Wishart (1)(16) to explain the developmental stages of the object concept suggests that the conceptual problem which underlies these stages for young infants is one of object *identity* rather than object *permanence*. A basic idea of object reality (including some idea of permanence) is assumed to be present in infants from birth (1); the infant is seen as having difficulty in maintaining the identity of an object throughout an event sequence. This difficulty is present whenever the event entails temporary disappearance of the object (4), and is particularly acute if the sequence involves close interaction with any other object (16).

In this theory, development is seen as a progressive refinement of the infant's rules for attributing identity to an object over time. The infant moves from the simple recognition that an object is the same object at different times and in different places, through to more elaborate notions which define identity in a much stricter sense, with the object not only being recognized as perceptually the same but as identical in the sense of being one and the same object when involved in any event sequence, i.e., the same and only such object involved.

Each change in development level means that the infant can maintain the identity of an object over increasingly complex event sequences. Each new identity rule reduces the population of "objects" with which the infant must deal and therefore represents a considerable cognitive achievement.

These rules and the psychological evidence for their validity are outlined below (for a fuller account, see (16)).

Rule 1, which corresponds to Piaget's stages I and II, (10) (11) is stated:

*An object is a bounded volume of space in a particular place or in a particular path of movement.*

It follows from this rule that two objects cannot be in the same place and that two objects cannot be on the same path of movement. A violation of rule 1, such as replacement of a stationary object by a totally different object, will be treated by this level of infant as a transformation of the original object rather than as a replacement by another object (1).

Application of this rule in search tasks would lead to the following search behaviors:

- to find a stationary object, look for it in the place where it usually is.
- to locate a moving object, look for it along its path of movement.

When an object that was stationary begins to move and the subject looks back to the space the object previously occupied what has become known as a *place error* occurs (2). When a moving object has in fact stopped and the subject continues to follow its path, a *movement error* occurs (3).

Rule 2 describes the second stage of development and corresponds to Piaget's stages III-V (10)(11):

*An object is a bounded volume of space of a certain size, shape, and color which can move from place to place along trajectories.*

Place and movement errors no longer take place because they are mediated by the perceptual features of the object, which were ignored in the application of rule 1. It is still true that two objects cannot be in the same place

\* We wish to thank the British Medical Research Council and North Atlantic Treaty Organization for supporting this research

\*\* Department of Computer Science, University of New Mexico, Albuquerque NM, USA

\*\*\* Department of Psychology, University of Edinburgh, Edinburgh, Scotland, UK

or on the same path of movement at the same time, i.e. that the *bounded volume* of space that *defines* the object cannot be violated. Total or partial occlusion of the object still causes problems for infants operating with rule 2.

Search behavior for this level of infant will include finding an object by searching for it its usual place, or if it has moved, along its usual path of movement. Since perceived information is incorporated in this rule for identifying an object, any sequence of actions violating the perceptual integrity of the object, as when the object is covered by a cup or some other occluder, will be treated by the infant as the replacement of the object by the new object. Behavior in this situation will be:

- to find an object that has mysteriously disappeared, remove the object replacing it, and with experience,
- to find the disappearing object, remove the object which is in the place where the desired object was last seen.

This will allow the infant to succeed in Piaget's stage III-IV or IV-V tasks, but does not represent any true understanding of spatial interactions between objects.

Rule 3 corresponds to Piaget's Stage VI (10)(11):

*Two or more objects cannot be in the same place or on the same path of movement simultaneously unless they bear a spatial relationship to each other which involves the sharing of common boundaries.*

The third identity rule is essentially the same as rule 2 but is modified to fit with the infant's experiences of the consequences of interactions between objects. Thus, for an infant working with only rule 1, an object which moves then stops or which enters into a spatial relationship with another object in such a way as to lose or mask its identifying boundaries will have disappeared mysteriously. With rule 2, only the latter kind of event will cause identity confusion and erroneous search behavior. Not until acquisition of rule 3 can the infant understand that a spatial relationship between two objects does not violate the identity of either. Prior to the acquisition of rule 3, the infant does not understand that a spatial relationship between two objects does not violate the identity of either.

In summary then, Bower and Wishart hypothesize that the infant develops a progressively more comprehensive set of rules for recognizing and maintaining the identity of an object over time. The staged acquisition of these rules both directs the infant's attempts to relocate objects and explains the erroneous behavior seen on the traditional object permanence tests. As one rule is replaced by the next, the infant comes closer to the appreciation of the invariant properties of individual objects. At maturity these rules will be sufficiently developed to allow an object to interact in common space with any other object without risk to its unique identity.

A first attempt at computer simulation of this period of cognitive development has already been made and is reported in (8). That paper deals only with the first two of the three hypothesised stages of development, comparing the output of the computer model with data collected in five Bower et al studies of infants' responses to simple object movements. The present paper describes the modeling of all three stages of development in two experiments.

A description of the PROLOG (6)(15) rules modeling each stage of development is now given. Rather than PROLOG code we list the competencies, expectations, etc. that make up the rules for each stage of development.

#### Stage 1.

- a. Focus on a location. This location has been *constructed* from the locations of the immediately preceding object structures found (see (e) below).
- b. Find an object within a fixed distance of where focused. If an object cannot be found report failure and look back to the preceding object found (the previous *snapshot*).
- c. Check the object for interest, seeing if it has volume or mass. This is done by considering two slightly different views of the object.
- d. Check if all boundaries are intact. This is done by checking the integrity of the boundary within the snapshot.
- e. Based on the object at snapshot (n) and snapshot (n - 1) construct an appropriate expected location for snapshot (n + 1).

#### Stage 2.

The competencies and expectations of stage 2 are almost identical to those of stage 1, as one might expect, except that a check occurs between (b) and (c) above comparing further perceptual relationships (size, color, shape) between the object at snapshot (n - 1) and the object found at snapshot (n).

#### Stage 3.

The competencies of stage 3 include all those at stage 2 and in addition the perceptual check of stage 2 is made both between (b) and (c), as in stage 2, and after the boundaries are scrutinized, after (d) above.

Experiments are run by our program in two independent steps: the creation of the set of snapshots of object structures that represent the physical experimental situation and then the analysis of these object structures by the rules of a particular developmental stage.

Our two steps, creation and analysis, provide for (in fact have an a priori commitment to) an independence of the object structure and its perception. This means that there is no interaction between the percept of an object and the subject that in any way changes the nature of the percept. The changes come in the subject's interpretation of that percept. This commitment to the primacy of perception allows description of its origins and presence according to a number of differing theories (6)(13).

## II. The Experiments and Discussion

Two experiments are described in this paper. All three stages of the model are tested on each of these experiments. The snapshots for the experiments are described in Figures 1 and 2.

In experiment 1 (Figure 1) a yellow sphere of radius 4 is located at (60,4,10) at time 1. (60,4,10) marks the (x - left/right, y - up/down, z - depth) coordinates of a 3-dimensional cartesian space. The sphere remains stationary for three time periods (or snapshots) and then moves to the right (with respect to the infant who views the scene from the point (60,0,0)). After moving for three time periods the sphere arrives at (72,4,10) where it again rests for three time periods before moving left and back to the starting point (60,4,10). This same sequence of rest and motion is repeated three more times. Then, instead of moving off to the right as usual, the sphere moves to the left for three time periods, coming to rest at location (48,8,10). Here the experiment ends.

Experiment 2 (Figure 2) has two objects, a green cube and an occluder which the infant views from position (36, 0, 0). The occluder, in this case a black platform of length 8, height 6 and depth 6, remains centered at location (36, 4, 6) for all time periods. (A platform is not conventionally thought to be an occluder. However infants produce the same sequence of behaviors with platforms as they do with more traditional occluders such as screens or tunnels (10)(16)). The green cube of length 4, height 4, and depth 4 remains at location (4, 8, 10) for the first five time periods. From times 6 to 20 it moves right, with respect to the infant, until it rests at location (68, 8, 10). While passing to the right the bottom boundary of the cube is obscured by the platform from time 11 to 15. After resting at (68, 8, 10), the object retraces the path back to the original starting place (4, 8, 10), repeating the partial occlusion in the middle of the path.

Each experiment of this study was chosen for a reason. Experiment 1 was an experiment Prazdny (12), in his computational description of certain Bower and Wishart experiments was unable to model. Experiment 2 of this study is an experiment already run with human subjects with known outcomes. In fact, results very similar to those described below have been reported by Bower and Wishart (4)(16). The second experiment is also important in current work at Edinburgh designing a cost/gain model of development (see below).

The results:

Experiment 1: The PROLOG model for stage 1 produced movement and place errors each time the object either stopped or started in motion, 15 times altogether. There was no problem following the new motion in a different direction as long as the object's locations were close enough to each other across consecutive time intervals. ("Close enough" is an empirically testable measure with infants). Stage 2 infants perceived only one object since their perceptual checks were able to determine two objects to be one and the same if color and size measures remained invariant across time. Because there were no border violations stage 3 analysis gave the same results as stage 2.

Experiment 2: Stage 1 found a new object when either motion or rest or boundedness was violated (9 in all). Stage 2, using perceptual checks, found new objects only when boundedness was violated (5 objects). Stage 3 found only one object since perceptual checks were able to override violations of boundedness.

A motivating force in the design of our model has been to demonstrate that the infant's detection of the presence or absence of different perceptual invariants across the object structures that make up the experiments offer two powerful explanatory mechanisms: First, we hypothesize that the psychological effect of the presence of such formal (i.e. internal or mental) invariants is to produce the behaviors that allow us to discern the three distinct stages of development. The infants' tracking behavior is directed not by *direct* perceptual input (14) but by three sets of conceptual rules discerning the invariants found across the object structures of the experiments.

Second, we hypothesize that the high cost of coping with multiple *objects* (non integrated perceptual phenomena) at one stage of development gives way to a more economical accounting with the discovery of new perceptual invariants at the next stage. The discovery of new invariants provides a more parsimonious explanation for the same phenomena. This cost/gain explanation for between stage development is the main focus of our continuing research, with the model and with real babies. At present we know that infants exposed to tasks like experiment 2 described above show accelerated development through the stages. The acceleration is manifested not only in simple, visual tracking tasks such as this experiment but also in transfer tasks involving manual search (16).

The fact of acceleration in the transfer tasks gives us confidence that the changes induced by tracking experience are conceptual changes rather than changes in sensori-motor skill. Our hypothesis is that the motive force of



change is the conceptual gain associated with the reduction in the number of phenomena the cognitive system must deal with at each successive stage. The 'cost' of this 'gain' is the increased load the perceptual system must bear.

With both experiments there is a clear gain: Experiment 1 presents an object in motion and at rest and then changing direction. The youngest infants perceive a new object each time the rest or motion change. These multiple objects are unified by the perception of color and shape invariance at stage 2. Similarly, in experiment 2 multiple phenomena are perceived as the object starts, stops, and shares boundaries with the platform. Again the invariant aspects of the object, even as it shares a common space, allow the more advanced infant to perceive a single object at rest, moving, and sharing boundaries.

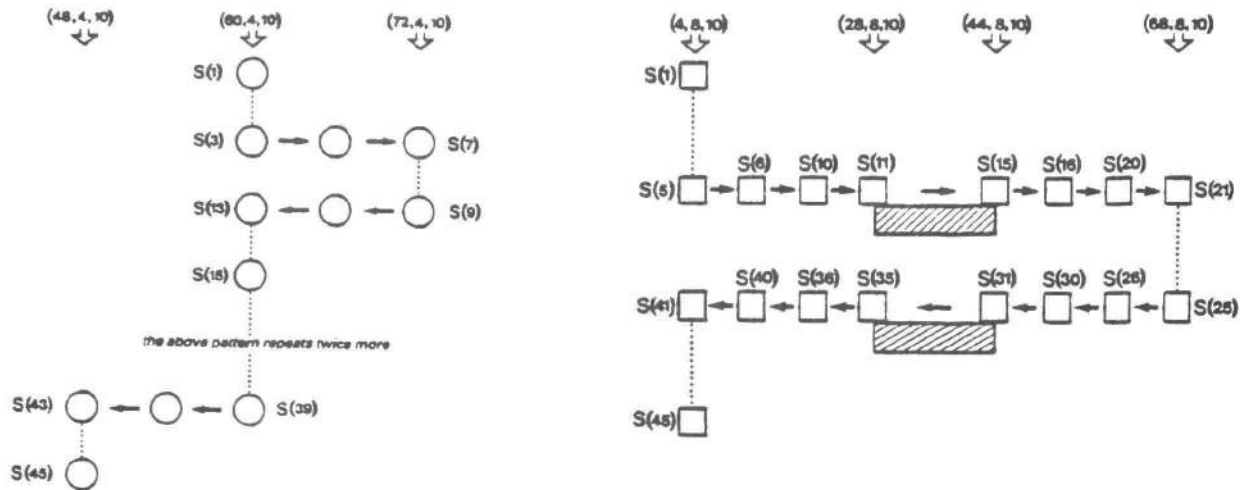


Figure 1 (l) and Figure 2 (r). S(i) represents the i-th snapshot or time period for the experiment and (x,y,z) describes the location of the object in 3-dimensional space.

## References

- (1) Bower, T.G.R., 1974. *Development in Infancy*. San Francisco, W.H. Freeman.
- (2) Bower, T.G.R., Broughton, J.M., and Moore, M.K., 1971. Development of the object concept as manifested in the changes in the tracking behavior of infants between 7 and 20 weeks. *Journal of Experimental Child Psychology*, 11, p 182-193.
- (3) Bower, T.G.R., and Paterson, J.G., 1973. The separation of place, movement, and time in the world of the infant. *Journal of Experimental Child Psychology*, 15, p 161-168.
- (4) Bower, T.G.R., and Wishart, J.G., 1973. The effects of motor-skill on object permanence. *Cognition*, 1, pp 165-171.
- (5) Bresson, F., Maury, L., Le Bonniec, G.P. and de Schonen, S., 1977. Organization and Lateralization of reaching in infants: an instance of asymmetric functions in hands collaboration, *Neuropsychologia*, 15, pp 311-320.
- (6) Clocksin, W. and Mellish, C., 1981. *Programming in PROLOG*. Springer-Verlag.
- (7) Luger, G.F., 1981. Mathematical model building in the solution of mechanics problems: human protocols and the MECHO trace. *Cognitive Science*, 5, 55-77.
- (8) Luger, G.F., Bower, T.G.R. and Wishart, J.G., 1982. A model of development of the early infant object concept. *Perception* (in press).
- (9) Marr, D., 1978. Representing visual information. In: *Computer Vision Systems* (eds. Hanson and Riseman) New York, Academic Press.
- (10) Piaget, J., 1936. *The Origins of Intelligence in Children*. London, Routledge and Kegan Paul, 1953.
- (11) Piaget, J., 1937. *The Construction of Reality in the Child*. London, Routledge and Kegan Paul, 1954.
- (12) Prazdny, S., 1980. A computational study of a period of infant object-concept development. *Perception*, 9, p 125-150.
- (13) Ullman, S., 1978. *The Interpretation of Visual Motion*. Cambridge, Mass., MIT Press.
- (14) Ullman, S., 1980. *Against Direct Perception*, AI memo 574, Artificial Intelligence Laboratory, MIT.
- (15) Warren, D., and Periera, 1977. PROLOG - The Language and its Implementation Compared with LISP, *ACM SIGPLAN Notice*, 12(8) and *SIGART Newsletter*, No. 64, 109-115.
- (16) Wishart, J.G., 1979. *The Development of the Object Concept in Infancy*. Unpublished Doctoral Dissertation, Department of Psychology, University of Edinburgh.
- (17) Wishart, J.G. and Bower, T.G.R., 1982. Spatial relations and the object concept: a normative study. In: *Advances in Infancy Research*, Vol. 3, (ed. L.P. Lipsitt and C. Rovee-Collier) Norwood, N.J.: Ablex.

# **Paper Session #5**



# Reasoning about the Temporal Structure of Narrative Texts

by

Michael J. Almeida and Stuart C. Shapiro  
Department of Computer Science  
State University of New York at Buffalo  
4226 Ridge Lea Road, Amherst, New York 14226

## Abstract

The process of determining the temporal structure of a narrative text is extremely complex. In this paper, we examine a small but central part of this process: the roles of aspectual class and the progressive/non-progressive distinction. A standard set of aspectual classes is presented and the temporal effects of each of these classes is discussed. Finally, we briefly discuss an implementation of a system which can read a simple narrative text and construct a model of the temporal structure of that text.

## 1. Introduction.

We are investigating the process of determining the temporal structure of narrative texts. This entails an investigation into the many factors which operate together within a narrative to indicate the temporal relations which hold between the events and situations mentioned in the text. Among these factors are tense, the progressive/non-progressive distinction, time-adverbials, world-knowledge, and aspectual class. In this paper, we will examine the roles of aspectual class and the progressive/nonprogressive distinction in this process.

We are using an essentially interval-based approach (similar to that proposed in Allen [1]) to the representation of temporal information. We do, however, use some time-points. Whether these are "really" points or just very small intervals is a question which we do not address.

## 2. The Narrative Now-Point.

Within a narrative, the most important temporal reference point is the point which represents the "present" moment within the narrative. When the time adverbial now is used in a narrative, it is this point of time which is being referred to. For this reason, we refer to this reference point as the narrative now-point, or, more briefly, as the now-point.

This now-point functions within a narrative more or less the way that the present (the "real" now) functions in the real world. That is, everything which comes before the now-point is in the past (in the world of the story) and everything that comes after the now-point is in the



future from the perspective of that moment in the story. Thus, as the story progresses in time, the now-point is moved forward in time. Uncovering, and then modeling, the ways in which different temporal devices interact with, and sometimes affect, this now-point is one of the major goals of our research.

### 3. Aspectual Class.

The notion of aspectual class plays a central role in the determination of the temporal structure of a narrative. One of the earliest detailed discussions of the concept of aspectual class was given in Vendler [7]. This work has since been refined and expanded upon in Dowty [2], Vlach [8], and Steedman [6], among others. The basic idea is that propositions can be shown to fall into one of a small number of categories (sometimes called aspectual classes) based on a number of properties of these propositions. The exact number of such classes, along with the catalog of properties associated with each class, varies with the researcher. Vendler distinguished the following four:

#### 3.1. Achievements.

Examples of this class are: "Mary fell asleep at 12 o'clock", and "John reached the top of the hill". The principle temporal properties of this class are (1) the simple, i.e. non-progressive, forms are true only at time points; and (2) the truth of the progressive form (if it exists at all) does not imply the truth of the simple form, so that for instance, "John was falling asleep" does not necessarily imply that "John fell asleep".

#### 3.2 Accomplishments.

Examples of this class are: "John ran a mile" and "Mary played a sonata". Typically, accomplishments involve a goal or an outcome of some sort. The principle temporal properties of this class are: (1) the simple form is true at an interval of time, i.e. accomplishments are not point-like; and (2) like achievements, the truth of the progressive form does not imply the truth of the simple form, so that for instance, "Mary was painting a picture" does not mean that "Mary painted a picture", i.e. that she completed it. Dowty [2] refers to this property of both accomplishments and achievements as the "imperfective paradox". That we are aware of, the first person within A.I. who dealt with this problem was McDermott [3]. He used an "in-progress" operator to mark the progressive forms. Vlach [8] also argues for the use of a PROCESS operator and this is the approach that we are also using. On the other hand, the simple form of an accomplishment does imply the progressive form, so that if "John ran a mile" then "John was running a mile".

The class of propositions consisting of the union of the achievements and the accomplishments (both in their simple forms) has been referred to as the class of events [4]. The way events typically behave can be seen in the following example. Imagine we are reading a text and we come across the following sequence of sentences:

John got out of bed. He wrapped his blanket about  
his head and shoulders.

The meaning of this particular example seems to be that first John got out of bed and then he wrapped the blanket about his head and shoulders. We interpret these sentences this way because (1) we know we are reading a narrative; (2) the sentences appear in a particular order; and (3) these sentences describe events. Notice that world-knowledge is not helpful in this example. Typically, an event sentence in the simple past tense describes an event which occurs AFTER the previously established now-point, and it has the effect of updating this now-point to just AFTER the time-interval of the event described.

### 3.3. States.

Examples of states are: "John knew the answer" and "The jar was on the table". The principle temporal properties of states are: (1) if a state is true for some interval of time then it is true for all points and subintervals of that interval; and (2) states are always viewed imperfectively, i.e., from within.

As a demonstration of the effect of the second property, consider the following example which consists of an event sentence followed by a stative:

John awoke. It was dark in the room.

The most likely interpretation of this example is that the state of its being dark held not only for some interval after John awoke but also for some interval before he awoke. What we believe happens in such an example is that the event (John awoke) establishes a new now-point in the usual manner, and then the state is viewed imperfectively from that now-point, that is, the now-point is DURING the time-interval associated with the state. Typically, a state does not move the now-point. If we were to add one or more additional stative sentences to this example, then all of these states would relate in this same way to this one now-point. This sort of "piling on" of states is common in descriptive sections of narratives.

Notice that the belief that the state actually overlaps the preceding event does not directly follow from the mere fact that the state is viewed imperfectively from the now-point, but is in fact an inference that we would probably want the system to make in this case. In the majority of such cases this seems to be a proper inference to make.

As a case in which this inference does not go through, consider the following example:

John turned off the light. It was dark in the room.

In this case the state is still viewed imperfectively from the now-point established by the event, but this time, the system should infer that the state of its being dark has a start-time after the event's end-time, but of course, still before the now-point. And so the state does not overlap the event which is apparently its cause.

In both cases, the basic way in which states behave with respect to the now-point is the same, but the system can then make additional inferences based on world-knowledge (assuming that it has this

knowledge) which further decide whether or not there is overlap.

### 3.4. Activities.

The fourth and last category of propositions distinguished by Vendler is the class of activities. Examples of activities are: "John listened to music", and "Mary played the piano". Unlike accomplishments, activities do not have an intrinsic goal or expected outcome.

The principal temporal properties of activities are: (1) they are true at intervals of time, i.e., they are not point-like; and (2) the truth of the progressive form of an activity does imply the truth of the simple form, so that, for instance, "John was listening to music (but was interrupted)" allows us to infer that "John listened to music". Also, of course, the simple form does imply the progressive form.

In a narrative, simple activities behave in a way which is intermediate between that of events and states. For example in:

John walked into the office. The secretary typed at her desk.

The secretary's typing (an activity) possibly overlaps the preceding event, and in addition, may still be continuing. However, unlike states, simple activities cause the now-point to move forward. Thus, successive simple activities may or may not overlap each other.

### 4. The Progressive.

Vlach [8] proposes that the function of the progressive operator is to change non-statives into statives. Certainly the temporal behavior of progressives is in many (but probably not all) circumstances the same as that of states. In the absence of time adverbials, the progressive forms of the three non-stative aspectual classes behave with respect to the now-point in exactly the same way as states do.

### 5. An Implementation.

We have implemented a system which can read a simple narrative text and construct a model of the temporal structure of that text. The system works as follows:

1) a sentence is parsed, a representation of the tenseless proposition is built, and the proposition's aspectual class is determined.

2) depending on the aspectual class and whether the sentence is simple or progressive, the proposition is related to its associated time-interval or point by one or both of the following case frames: (1) PROPTIME-PROP : which means that the simple form of the proposition is true at (or in the case of states, for) that interval; (2) PROGTIME-PROG : which means that the progressive form of the proposition is true for that interval.

3) this time-interval (or point) is then related to the current now-point in one of the manners described earlier in this paper, and then finally,

4) in the case of events and simple activities, the now-point is updated to a position AFTER the just added time-interval.

5) the next sentence is now read.

The parser is implemented as an ATN and the representations are in the form of semantic networks implemented in the SNePS semantic network processing system [5]. The current implementation will only accept sentences in the simple past or the past progressive tense. In addition, the system does not handle time adverbials or do any inferencing. We expect to expand and improve the system as our research progresses.

## 6. Conclusion.

We have discussed some of the major temporal properties of the various aspectual classes and we have indicated how, in conjunction with the progressive/non-progressive distinction, we believe these classes behave in a narrative text. We have also briefly discussed our implementation of these ideas.

## 7. References.

1. Allen J.F., Maintaining knowledge about temporal intervals, TR 86, Dept. of Computer Science, University of Rochester, 1981.
2. Dowty D.R., Word Meaning and Montague Grammar, D. Reidel Publishing Co., 1979.
3. McDermott D., A temporal logic for reasoning about processes and plans, Cognitive Science, Vol.6, pp.101-155, 1982.
4. Mourelatos A.P.D., Events, processes, and states, In P.J. Tedeschi and A. Zaenen (eds.), Tense and Aspect, Vol.14 of Syntax and Semantics, Academic Press, pp.191-212, 1981.
5. Shapiro S.C., The SNePS semantic network processing system, In N.V. Findler (ed.), Associative Networks: The Representation and Use of Knowledge by Computers, Academic Press, pp.179-203, 1979.
6. Steedman M.J., Verbs, time, and modality, Cognitive Science, Vol.1, pp.216-234, 1977.
7. Vendler Z., Verbs and times, In Z. Vendler, Linguistics in Philosophy, Cornell University Press, 1967.
8. Vlach F., The semantics of the progressive, In P.J. Tedeschi and A. Zaenen (eds.), Tense and Aspect, Vol.14 of Syntax and Semantics, Academic Press, pp.271-292, 1981.





# Event Concept Coherence in Narrative Text

Richard Alterman  
Department of Computer Science  
University of Texas at Austin  
Austin, Tx 78712

## What It Is.

For the purposes of this paper we will take the meaning of a concept to be its position in a structured network of mutually defining discrete symbols. For example, the concept 'give' is defined by its relationships to concepts like 'have' (i.e. to 'give' one must first 'have' the thing given, as a result of 'giving' there is a new 'having'), 'exchange' (i.e. 'giving' is a kind of 'exchange'), 'trading' (i.e. 'giving' and 'trading' are both kinds of 'exchange'), etc.

'Event/state concepts' are taken to be the words (terms) in a language which designate events or states. Two event/state terms, as they appear in text, are 'concept coherent' if their corresponding positions in a structured net of meaning are *proximal*. That is, given the text, "John gave Mary the bicycle. She has it.", the two event/state descriptions used in the text are 'concept coherent' because of the close inter-relationship between the concepts 'giving' and 'having'.

The point of an 'event concept coherence' representation of text has to do with the difficulty of establishing meaning representations. The idea is to capture only some of the properties of the text's meaning, to organize the text for further interpretation without losing any of the information of its original form. 'Event concept coherence' allows the text to be structured not so much on the basis of its meaning, but on the basis of a *property* of its meaning, its concept coherence. Because the proximity of concepts (i.e. concept coherence) is in some ways independent of the complexity of the relationships among terms in the network, it is possible to relax, at first, the content of the network and represent the connectivity of event/state concepts in terms of relationships easier to explain and establish. I will argue that initial structure for event concepts should be temporal relations, saving the causal, affect, and speech act interpretations for later stages.

The analysis of text in terms of event/state concept coherence results in a grouping of text into a hierarchy of chunks, each chunk centered around some event/state concept. Consider the following text adapted from "The Tale of the Pig" (Protter, 1961):

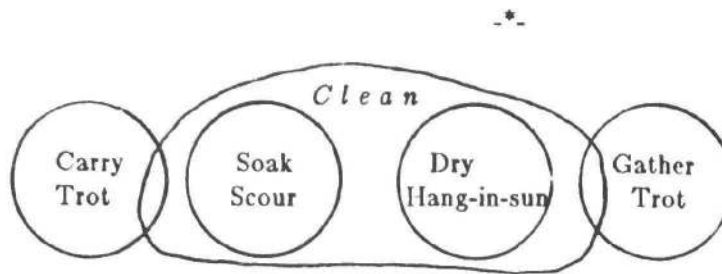
The pig trotted towards the stream carrying a bundle of clothes. The animal expertly soaked and scoured the laundry. The pig hung the clothes in the sun to dry. The pig gathered her laundry and trotted home.

The first set of terms which seem to cohere are 'trotting' and 'carrying'; 'carrying' involves 'travelling', which involves 'moving', and a kind of 'moving' is 'trotting'. The 'gathering' and 'trotting' described in the last sentence also seem to group around the concept 'carrying'; to 'carry' one must 'have' that which is carried, and 'gathering' results in a 'having'. 'Hanging clothes in the sun' is a way of 'drying', which, along with 'soaking', and 'scouring' are all parts of 'cleaning'. Finally, the 'cleaning' intersects the two 'carrying's' (i.e. the pig is 'moving' the laundry to the stream to 'clean' it and home again after it is 'cleaned'), so the text can be collected into a single chunk centered around the concept 'cleaning' (see figure 1).

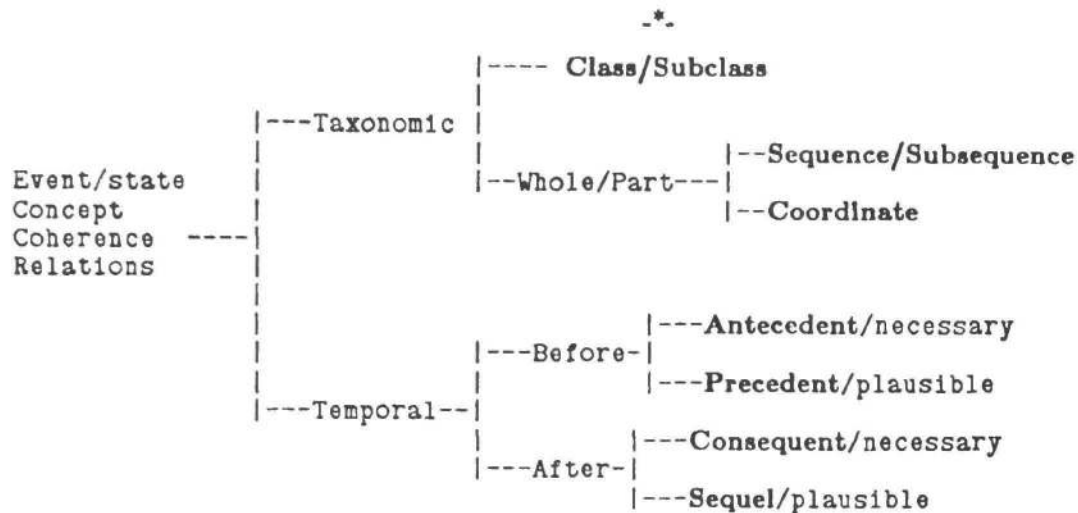
## A Little Structure

The ability to establish 'concept coherence' between event descriptions in text is dependent on the availability of a 'dictionary' (i.e. a structured network of concepts). In fact, an explanation (representation) of the concept coherence of the event descriptions in a piece of text can be derived by copying the relevant portion of the dictionary's structure. Because the dictionary is used for an initial structuring of the text, ideally its structure is simple, general, and useful.

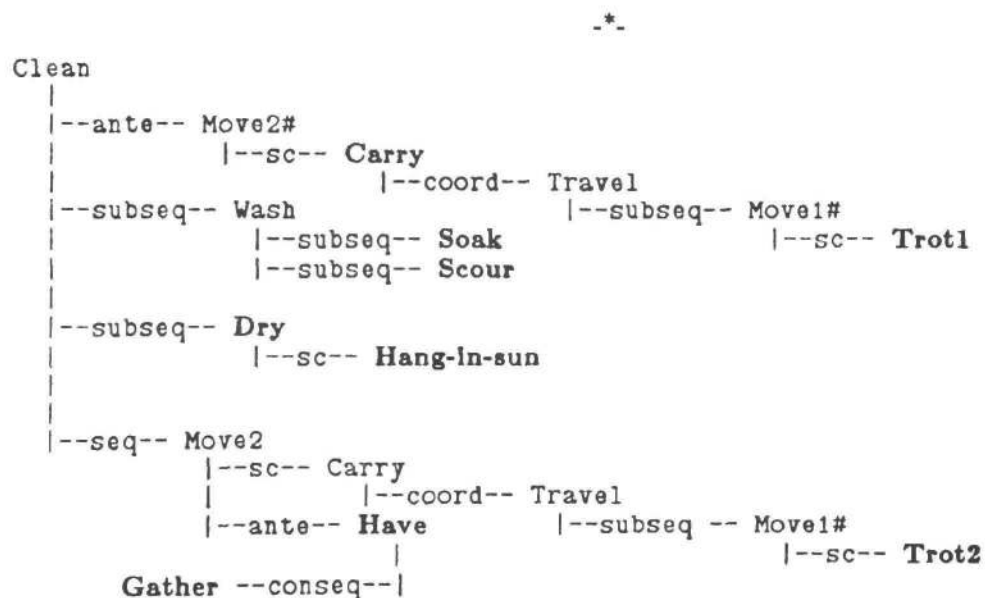
NEXUS, the system developed to test this theory, uses a dictionary of between 100 and 150 event/state concepts interrelated by seven concept coherence relations (see figure 2). There are three taxonomic relations; one is a property inheritance relation, and the other two whole/part relations. The taxonomic relations are used to organize individual concepts into a hierarchical structure. Four of the relations are temporal; these are used to chain together sequences of typically co-occurring events and states. *Class/subclass* (sc) is the property inheritance coherence relation. In text, to find a connection between a pair of event/state terms it is frequently necessary to infer the inherited properties of either or



**Figure 1:** The pig cleans the laundry.



**Figure 2:** Characterizing the Relations



**Figure 3:** The Tale of the Pig Revisited

-\*-



both terms. Consider the pair of sentences: "John has the tickets. He snatched them from Bill." There exists a direct relationship between 'having' and 'taking' (i.e. as a result of 'taking something' one 'has it'). Because of the class/subclass relationship between 'taking' and 'snatching' NEXUS is able to access this relationship and represent the coherence of the sentences.

Experiments with NEXUS showed that class/subclass was the most frequently occurring coherence relation. This is not surprising considering the economy of using subclass terms in text; a subclass term simultaneously refines the meaning of its parent class while maintaining a implicit access to all its coherence relationships.

The *sequence/subsequence* (subseq) and *coordinate* (coord) relations can be differentiated by their temporal intervals. If one event is a part of another event, and it occurs for a subinterval of time, then the corresponding concepts are in a *sequence/subsequence* relationship. Consider the event descriptions: "John farms an acre of land. He'll be planting in the spring." A sub-activity of 'farming' is 'planting', another is 'harvesting'. Whenever the concept 'farming' applies the concept 'planting' applies for a subinterval of the 'farming' event, consequently 'farming' and 'planting' have a *sequence/subsequence* relationship.

If an event has parts that co-occur over the same time interval, then the corresponding event concepts are in a *coordinate* relationship. For example, "John carried the book. He walked, holding it in his hands.", every event which conveys a sense of 'carrying' in part also conveys a sense of 'moving' and 'holding' throughout the duration of the carrying, and there are 'moving' and 'holding' events that are not part of a 'carrying', and no 'carrying' event can completely be described by 'moving' or 'holding' - thus 'moving' and 'holding' are coordinates of 'carrying'.

*Antecedent* (ante), *precedent* (prec), *consequent* (conseq), and *sequel* (seq) are all temporal coherence relations. There are two ways of splitting the temporal relations. Antecedent and precedent concepts come before an event, and consequent and sequel concepts come after an event. The temporal relations can also be divided into logical and plausible groups of relations. Antecedent and consequent related concepts are logically related. Precedent and sequel concepts are plausibly connected.

If one event must necessarily occur before another event, the relationship between their corresponding concepts is classified as *antecedent*. For example, "John had some food. He ate it.": to eat one must first 'have' the thing which is to be eaten. Thus an antecedent of 'eating' is 'having food'.

If one event always, necessarily, occurs immediately after the other, then the relationship between their corresponding concepts should be marked as *consequent*. Take the following example, "John gave Mary a red kite. Mary has a red kite." Events of 'giving' are necessarily immediately followed by states of 'having'.

If one event, with some regularity occurs before another event, the relationship between their corresponding concepts can be classified as *precedent*. For example, "John opened the door. He entered.", events which can be described as 'entering' are sometimes preceded by sequences which enable the 'entering' and can be described as 'opening'.

If one event follows another with some regularity, the relationship between their corresponding concepts is *sequel*. An example of a sequel relationship is: "John cleaned the laundry. He carried it home." Sequences of 'cleaning' are typically followed by sequences where the laundry is 'moved'.

Although causal relations would have better described the relationships between two concepts, NEXUS's dictionary uses temporal relations because they are easier to establish. Recall that earlier I said that by relaxing the content of the net some of the complexity of establishing an initial interpretation of the text would be mitigated. Consider the use of the concepts 'cleaning' and 'carrying': "The pig cleaned the laundry. She carried it home." The causal relationship between these two event descriptions is not at all clear; it is neither result, nor enablement or reason. Using temporal relations it is easier to describe the relationship; it is not necessary that after the pig cleans the laundry she carried it home, but it is highly likely (i.e. sequel). A similar case can be made for the text: "Wild Bill rode to town. He tied his horse to the hitching post in front of the saloon." The causal relationship between 'riding' and 'tying a horse to a hitching post' is not simple, but with temporal relations it is easy to call the relationship sequel.

Furthermore, because temporal relations are causally neutral some of the problems associated with context can be finessed. Consider an example discussed by Wilks (Wilks, 1977) (p244): "The rock fell off the cliff and crushed John's lunch. Peter pushed it." The chain from 'pushing' to 'crushing' can be interpreted in one of two ways as either causal or goal. If it is later learned that Peter was angry at John then the chain is a goal chain. But if, instead, it is learned that John was clearing a spot to pitch his tent, the chain is a causal one. A system which produces causal chains would commit to one interpretation or the other (in Wilks' case the goal one), and potentially have to backtrack. NEXUS, because its dictionary uses temporal relations can produce a representation. At a later stage, NEXUS'

initial structuring of the text can be used as a basis for constructing a causal interpretation.

## The Tale of the Pig Revisited.

Let us return to the example from "The Tale of the Pig", in this case adding some structures which a concept coherence representation would inherit from NEXUS' dictionary (see figure 3).

Consider the first 'carrying' event. Coordinates of 'carrying' are 'holding' and 'travelling'. Subsequences of 'travelling' are 'departing', 'moving', and 'arriving'. A subclass of 'moving' is 'trotting'. So NEXUS finds in its dictionary the path from 'carry' to 'trot' via the concepts 'travel' and 'move'.

Similarly the second 'carrying' event can be constructed. In this case, to connect 'carrying' to 'gathering', the 'carrying' must inherit from a second sense of 'moving' (i.e. John moved the box) an antecedent relationship to 'have' which is a consequent of 'gathering'.

The subsequence relations can be used to collect together the parts of 'cleaning'; subsequences of 'cleaning' are 'washing' and 'drying'. The subsequence relation is also used to collect together two parts of 'washing'; 'soaking' and 'scouring' are two subsequences of 'washing'. The 'hanging in the sun' is connected to 'drying' via a subclass arc. Finally the first 'carrying' is connected to the 'cleaning' via a 'moving#2' by an antecedent relation, and the second 'moving#2' by a sequel arc.

There are a few noteworthy features of this representation. In itself the representation shows the events of the story are coherent; the events can be collected together under the concept 'cleaning'. So the representation is de facto evidence of the text's coherency.

Notice that the analysis has produced two interconnected trees. In general an analysis of a piece of text will produce several interconnecting trees (one per major concept). The top node of a tree represents a summary of a concept. In summarizing this text, NEXUS deletes the tree centered around 'gathering', because it is 'preparatory', and produces the summary "The pig cleaned the laundry at the stream".

In the course of deriving the representation NEXUS resolves several references. For example it determines that 'animal' refers to 'pig', and that 'laundry' refers to 'clothes'. It also infers elided information; e.g. it infers that the pig cleaned the laundry at the stream.

NEXUS uses the representation as a basis for suggesting answers to several types of questions. The subsequence relationships between 'washing' and

'cleaning' and 'drying' and 'cleaning' is used to answer the questions, "Why did the pig wash the laundry?" and "Why did the pig dry the laundry?", with the statement, "She was cleaning the laundry." The subsequence relationships between 'washing' and 'cleaning' and 'drying' and 'cleaning' are also used by NEXUS to answer an 'how' question; "How did the pig clean the laundry? By washing and drying it." The combination of consequent and precedent relationships, subject to further semantic checks, can be used to answer 'goal' questions. For the question "Why did the pig gather the laundry?", NEXUS uses the coherence relations to suggest the answer "So she could carry it home."

## Some Results and Conclusions

*Event concept coherence* should not be confused with the *discourse coherence* described by Hobbs (Hobbs, 1978). 'Event concept coherence' is primarily concerned with the coherence aspect of event/state concept definitions, and secondarily with how it is reflected in text. 'Discourse coherence' characterizes the text as discourse; it models the "coherent continuation moves" of the author or speaker (Hobbs, 1979) p68). 'Event concept coherence' represents the text by the dictionary relationships among the event concepts used in the text. With the exception of two *occasion* relations, enable and cause, 'discourse coherence' uses rhetorical relations, like example, parallel, elaboration and contrast. 'Event concept coherence' includes in its representation an explanation of the connection between two concepts, i.e. the inference path it found in the dictionary. 'Discourse coherence' does not. 'Event concept coherence' is derived in a data-driven fashion; 'discourse coherence' works from the top down looking for one of a set of rhetorical relations. 'Event concept coherence' has a dictionary of event concepts as its major source of knowledge; 'discourse coherence' requires a synthesis of semantic, factual, expert, stylistic, and grammatical knowledge. These comments are meant to clarify the many differences that exist between 'event concept coherence' and 'discourse coherence', not to detract from Hobbs' work.

The event concept coherence representation system is realized in a program called NEXUS. The subsystem which produces the representations is called TRACE, the question answerer QUEST, and the summarizer SUM. The system is programmed in procedural logic using HCPRVR (Chester, 1980a, Chester, 1980b).

Concept coherence analysis has been applied to text previously accounted for by scripts (Schank & Abelson, 1977, Cullingford, 1978), plans (Schank & Abelson, 1977, Wilensky, 1978, Wilensky, 1981), story trees (Rumelhart, 1975), schema-narrative trees (Simmons, 1982, Simmons, 1983, Correia, 1980),

and speech act (Allen, 1979, Cohen and Perrault, 1979, Cohen, Perrault, and Allen, 1981) theories. Event concept coherence is not intended to replace these top-down theories, but more to complement them with a theory that attempts to grow the more complex interpretations of text from the bottom-up.

TRACE has successfully been applied to eight samples of text. It is important to remember that for each of the eight examples TRACE used the identical dictionary of concepts. Three of the samples came from the AI literature; these included stories which had been accounted for by either scripts (Lehnert, 1977), plans (Wilensky, 1978) or story trees (Rumelhart, 1975). The other five samples, including a more difficult version of "The Tale of the Pig" example, come from a book of folktales (Protter, 1961). For further details see Alterman (Alterman, 1982).

## Acknowledgements

I would like to thank Bob Simmons and Elaine Rich for their comments on an earlier draft of this paper.

## References

- Allen, James. *A plan-based approach to speech act recognition*. PhD thesis, University of Toronto, 1979.
- Alterman, Richard. *A system of seven coherence relations for hierarchically organizing event concepts in text*. Technical Report TR-188, University of Texas at Austin, 1982.
- Chester, Dan. *Using HCPRVR*. Technical Report, The University of Texas at Austin, 1980.
- Chester, D. *HCPRVR: a logic program interpreter in LISP*. AAAI, 1980.
- Cohen, Phillip R. and Perrault, C. Raymond. Elements of a plan-based theory of speech acts. *Cognitive Science*, 1979, 3(4), 177-202.
- Cohen, P. R., Perrault, C. R., and Allen, J. F. *Beyond Question-Answering*. Technical Report, Bolt Beranek and Newman Inc., 1981.
- Correia, Alfred. Computing story trees. *American Journal of Computational Linguistics*, 1980, 6, 135-149.
- Cullingford. *Script application: computer understanding of newspaper stories*. PhD thesis, Yale University, 1978.
- Hobbs, Jerry. *Why is discourse coherent?* Technical Report Technical Note 176, SRI, 1978.
- Hobbs, Jerry. Coherence and coreference. *Cognitive Science*, 1979, 3(1), 67-90.
- Lehnert, Wendy. *The Process of Question Answering*. PhD thesis, Yale University, 1977.
- Protter, Eric (Ed.). *A children's treasury of folk and fairy tales*. : Channel Press Inc 1961.
- Rumelhart, David E. Notes on a schema for stories. In Bobrow, Daniel G. and Collins, Allan (Eds.), *Representation and Understanding*, : Academic Press, Inc, 1975.
- Schank, R. C. and Abelson, R. P. *Scripts, plans, goals, and understanding*. : Lawrence Erlbaum Associates 1977.
- Simmons, Robert F. A narrative schema in procedural logic. In Clark, K. L. and Tarnlund, S.-A. (Ed.), *Logic programming*, : Academic Press, 1982.
- Simmons, Robert F. *Computations from the english*. : Prentice-Hall 1983. In Press.
- Wilensky, R. *Understanding goal-based stories*. PhD thesis, Yale University, 1978.
- Wilensky, R. PAM. In (Ed.), Schank, R. C. and Riesbeck, Christopher (Eds.), *Inside computer understanding: five programs plus miniatures*, : Lawrence Erlbaum Associates, 1981.
- Wilks, Yorick. What sort of a taxonomy of causation to we need? *Cognitive Science*, 1977, 1, 235-264.



# Perseverers, Recencies and Deferrers: New experimental evidence for multiple inference strategies in understanding

Richard H. Granger  
Jennifer K. Holbrook

Artificial Intelligence Project  
Computer Science Department  
University of California  
Irvine, California 92717

In the course of understanding a text, a succession of decision points arise at which readers are faced with the task of choosing among alternative possible interpretations of that text. We present new experimental evidence that different readers use different *inference strategies* to guide their inference behavior during understanding. The choices available to an understander range from various alternative inferential paths to the option of making no inference at a particular point, leaving a 'loose end'. Different inference strategies result in observably different behaviors during understanding, including consistent differences in reading times, and different interpretations of a text. The preliminary experimental results given here so far consistently support a previously published set of hypotheses about the inference process that we have called Judgmental Inference theory.

## 1.0 Introduction

When trying to understand even a simple text, readers make complex evaluations of the text to help choose one of several alternative interpretations. In making such decisions, readers employ a number of different strategies, including the selection of an *inference path*. This paper presents a theory of how readers make such choices. For example, subjects in our experiments read the following story:

[1] Melissa began to cry. Tyler had just asked her to marry him.

When asked why Melissa began to cry, different subjects gave at least two significantly different answers: (1) because she was upset for some reason about Tyler's proposal, perhaps because she couldn't or didn't want to accept the proposal; vs. (2) because she was so happy about the proposal that she was crying tears of joy. Preliminary findings in a series of controlled experiments indicate that

1. the answers a subject gives to this and similar questions correlate with different reading times;
2. these differences in reading times, and the differences in question-answering behavior, and the correlation between them, are all accurately predicted on the basis of hypothesized 'inference strategies' contained in Judgmental Inference theory [Granger 1982];
3. the different reading times and interpretations are not due to arbitrary individual differences, but rather to the adoption by

individuals of some particular principled inference strategy, as evidenced by the fact that individuals initially exhibiting one type of strategy can be experimentally manipulated to exhibit a different type of strategy.

While many theories of inference in understanding acknowledge the existence of alternative inferential paths, we present a theory that (1) catalogs the paths available to an understander, (2) predicts what mechanisms will lead to the choice of particular inferential paths during understanding, and (3) provides preliminary experimental evidence which supports these hypotheses.

## 2.0 Background

Language understanding is an interactive process which requires both adequate information presentation on the part of the communicator and skills of interpretation on the part of an understander. The interpretation skills necessary for understanding include mechanisms for such inference tasks as associating referents (Anderson & Bower, 1973; Haviland & Clarke, 1974), recognizing temporally or causally related events (Schank & Abelson, 1977; Bower, Black, & Turner, 1979; Black and Bern, 1980), and filling in unstated actions and preconditions for actions in a stereotypical sequence of events (Schank, 1975, 1977; Bower et al., 1979).

For example, consider the following story:

[2] Gail had accidentally poured coffee onto Will's hand. He screamed bloody murder.

Readers have no trouble inferring that "he" in the second sentence refers to Will from the first sentence. This is known as *referential cohesion*. Readers also recognize that Will screamed after the coffee was poured on him, although no specific temporal connection was stated. Further, readers will only infer that Will screamed because he was in pain or because he was very angry with Gail (or both). They will not infer that he screamed because he saw a ghost, or because he suddenly remembered it was time for his primal scream therapy. This connection between the two events is known as *causal cohesion* (Schank, 1977). Causal cohesion has been demonstrated extensively through various experiments (Anderson et al, 1973; de Villiers, 1974; Bower et al, 1979; Black et al, 1980), which indicate that recall of story events was higher and reading time was faster if the events in the story were causally related. These studies demonstrate the integral role inferences play in understanding text.

It is important to note that understanders make inferences about text as they read, not after they have finished reading. Referential inferences have been demonstrated to be made during reading. Seifert, Robertson, and Black (1982) discuss Haviland and Clark's 1974 experiments which show that reading times are longer when statements cannot be easily connected through referents. Seifert



et. al. also point out that inferences which do not rely on linguistic context are made by the reader. These *pragmatic inferences* involve causal connections formed between statements in the text and world knowledge which a reader draws upon to infer causality, case relations, missing events in a stereotypical sequence of events, and other abstract inferences. For illustration, recall story [2]. At the end of the first sentence an inference is made that the subsequent events will be related to the coffee spilling. This is easily illustrated by giving a different version of the story to readers:

[2a] Gail had accidentally poured coffee onto Will's hand. He told her a silly joke.

Readers would be confused by this version because there is no apparent causal connection. They would try to relate the events in any way possible; for example, they might decide that Will told the joke to Gail before she poured the coffee, and that she found the joke so funny that she got careless with her coffee pouring. A reader might also decide that telling jokes was Will's stoic response to pain, or that the coffee wasn't hot. Most readers will eventually conclude that the events were not causally related at all. What is intriguing is that any original expectations about events which a reader generates after reading the first sentence will be different from interpretations possible after reading the second sentence. There must be at least two points during the reading at which inferences about events can be made, for instance, one after the first sentence, and one after the second. Such points are known as *inference points* (see Rumelhart, 1981).

Several options are available to the reader at an inference point. The reader may leave a *loose end* (Granger, 1980a, 1981); that is, no particular inference about further events is generated. The reader often comes up with default inferences, which, given the material already read, are the most likely of the possible events, reactions, and so on. In story [2], some of the default inferences would be that the coffee had scalded Will, that he was in pain, and that he might react to the pain. The reader can also make another kind of inference which would not be based on the most likely outcome of the story based on the events so far. For example, in story [2], a plausible inference is that Will may decide to take revenge on Gail, or that Will screamed before Gail spilled the coffee, and she spilled the coffee because the scream startled her. Intuitively, the default inference about temporal relations is that events are described in the order in which they happened. Likewise, the most frequent default inference about causal relations is that earlier events caused subsequent events to happen.

However, it is not always the case that the order in which events are relayed in text is the order in which the reader infers them to

have happened. It only appears this way because texts are often written in a problem/solution fashion, rather than an action/explanation fashion. There are many texts that do not make it immediately clear whether or not events have been relayed in cause/effect order. When this happens, different strategies may arrive at different interpretations of the order.

The inferences that are made conform to *evaluation metrics* (Granger, 1980a) which determine how appropriate the inference is. Cohesion, in its various forms, is one such evaluation metric. Another metric is *parsimony* (Granger, 1980a, 1980b). Parsimony refers to the observation that readers will infer the least complicated explanation of events possible. For example:

[3] Mary picked up a magazine. She swatted a fly.

In the above story, readers will always infer that Mary picked up the magazine in order to swat the fly, even though several other interpretations are possible (for example, that she picked up the magazine to read, then was annoyed by the fly and used the magazine to swat it. Readers will agree that these more complicated interpretations are equally possible interpretations. But even though these interpretations are all equally logically possible, people universally choose the most *parsimonious* of the interpretations -- i.e., the one that will explain most parsimoniously the maximum number of events in the text.

If the reader left a loose end after reading the first sentence in the story above, as in figure 1a, there would be no problem relating the two sentences with the most parsimonious inference. However, if the reader makes the default inference that Mary wanted to read the magazine, as in figure 1b, the reader must *supplant* that initial inference with the new inference that the magazine was going to be used for flyswatting. Supplanting occurs when a default inference made by a reader does not account parsimoniously for all of the events in the text.

Readers are constantly evaluating their own inferences, supplanting inferences which are not parsimonious, re-interpreting whole sections of text when the text has a surprise ending (e.g. mystery stories, jokes), and so on. Judgmental Inference theory [Granger 1982] hypothesizes that all inference points are in fact decision points, where one of many paths is chosen (e.g. leaving a loose end, making a default inference, supplanting an inference, etc.) based on complex evaluations of the interpretation being constructed. These interpretations can involve a series of intermediate steps, in which many interpretations are tested. Judgmental inference is not a conscious attempt to discover the interpretation that is most parsimonious, most logical, or most cohesive. Rather, it is an unconscious process, guided by inference strategies.

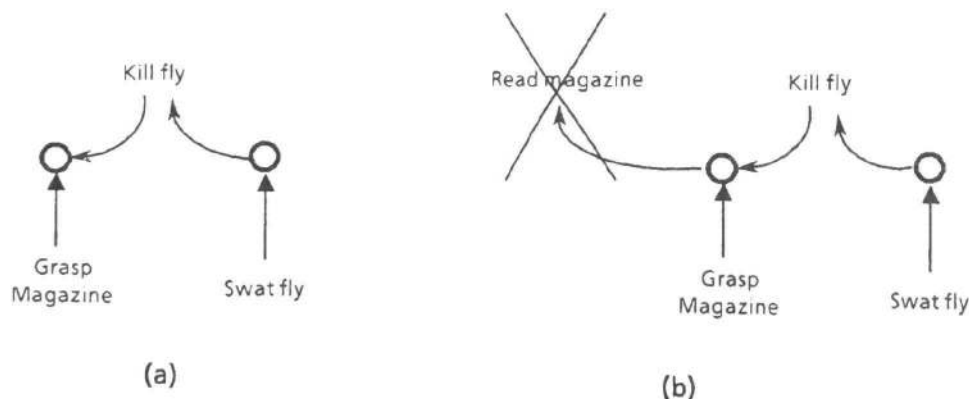


Figure 1



### 3.0 Theoretical Predictions of Judgmental Inference

Judgmental inference theory makes several specific predictions of how people choose 'inference paths'. This paper and the experiments described here focus on the study of these predictions. In particular, the experiments discussed below focus on the following questions:

1. When in the interpretation process are inferences made, if at all?
2. When in the interpretation process are loose ends left, if at all?
3. When do inferences get supplanted, if at all?
4. What happens when a reader doubts an inference?
5. When there are two (or more) equally plausible, parsimonious, and "normal" interpretations of story events possible, which one will readers select, why will they select it, and what will happen if doubt about the chosen interpretation is introduced?

It was pointed out above that a reader may take one of several inference paths when interpreting a story. For example, readers can leave loose ends or make inferences at inference points. Readers can read with a "naive" or a "suspicious" understanding: they can assume that the author is conveying the facts or trying to deceive them, that actors in the text have overt or covert goals, and so on. Yet most readers come up with very similar interpretations of story events. Either all readers follow the same inference paths, or with enough constraints, all inference paths will lead to similar interpretations of events. This theory espouses the latter view.

We theorize that interpretations are based on strategies or systematic choices between inference paths, and that individuals tend to use the same inference paths consistently. This would make possible the discovery of inference path choices, and explain why readers' interpretations of text are usually similar, yet diverge on occasion. However, the inference path system chosen is neither idiosyncratic nor universal to all readers. Instead, there seems to be a 'scale' of systems. We have divided this scale into several broad categories, although it should be noted that in reality, we have so far found no clear divisions among them.

Those readers who tend to come up with an interpretation of events as early as possible in the reading and then cling to that interpretation as long as possible can be described as *Perseverers*. Those readers who tend to leave loose ends about goals and plans (unless the goals or plans are explicitly stated) are known as

*Recencies*. Recencies will come up with an interpretation fairly late in the text, and if a conflict develops between possible interpretations, the interpretation based on the most recent information is chosen.

There are examples of extreme behavior at both ends of the scale. The most dedicated of Recencies will not make inferences. Unless a goal or a plan is explicitly stated, these readers will leave loose ends. Such behavior should result in quick reading, but slow and possibly haphazard answering when queried about inferred events. The extreme version of the Perseverer might be analogous to a "paranoid" reader. Such readers would make inferences based on preconceived notions. They might relate text to their own experiences, or ascribe attributes to characters which are clearly not in line with actual text.

Because there is not necessarily a clear division between the two main categories, there are readers who behave as though they could belong in either category. Such readers are called *Deferrers*. It is not clear whether Deferrers are using some combination of the other strategies or a different sort of strategy altogether. Future experiments may help explain Deferrer behavior.

Recall the following example:

- [1] Melissa began to cry. Tyler had just asked her to marry him.  
Q: Why did Melissa cry?

As Figure 2a shows, a Perseverer reading this story will believe that Melissa was upset with Tyler's proposal; maybe she doesn't like Tyler, or maybe she is unable to get married even though she loves Tyler. The exact cause is unknown, but the default inference based on her tears is that she is unhappy. That is, in the absence of more specific information, crying is assumed to be a visible sign of pain or unhappiness, and there is some action which can explain unhappiness. Any new information in the story (here, Tyler's proposal) will be interpreted as an explanation for that unhappiness or a reaction to the unhappiness.

A Recency would believe just the opposite: that Melissa is happy with the proposal, and that she is crying tears of joy, not tears of sorrow. Figure 2b illustrates the inference strategy of a Recency trying to understand the events in this story. Recencies do not make an initial inference about Melissa's crying, not even a default inference; unless a specific goal or plan is stated, a Recency will leave a *loose end*, waiting for more specific information. If later events in the story are more specific about goals and plans, then earlier events will be interpreted in light of this more specific information. Even if no more specific information is given, a

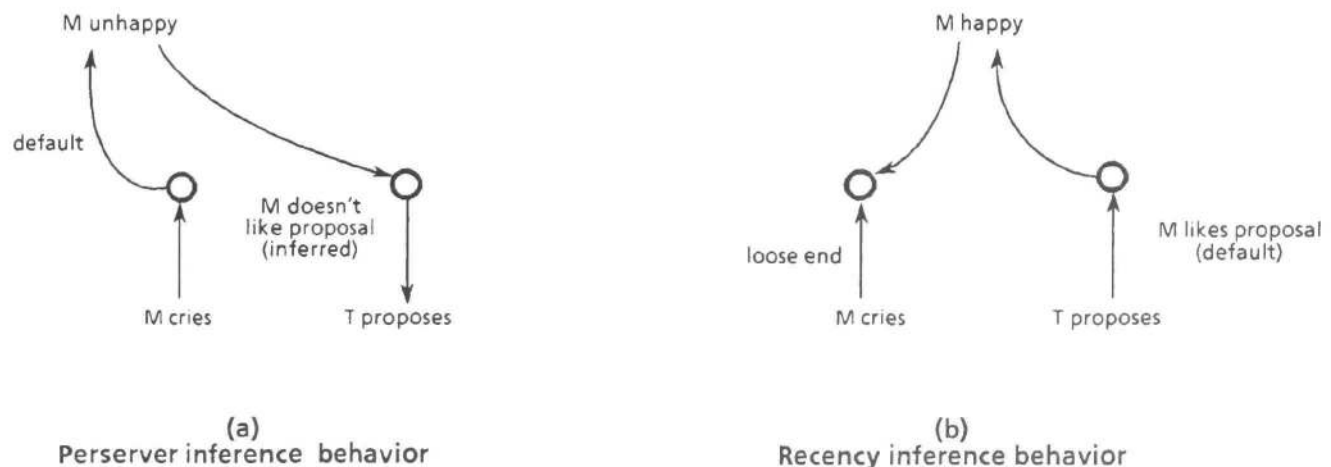


Figure 2

Recency will use the latest events in the story to interpret earlier events. Thus, a Recency would make no inference about the cause of Melissa's tears. The later information about Tyler's proposal would give rise to a presumed default interpretation of "happy event", and the earlier events in the story would be interpreted as being in line with a happy event, so that Melissa's tears are assumed to be tears of joy.

If this model is correct, reversal of the sentences of the above story should produce the opposite interpretation from the two extreme groups; that is, when presented with:

[4] Tyler had just asked Melissa to marry him. She began to cry.

Recencies should infer that Melissa is unhappy about the proposal, while Perseverers should infer that she is crying tears of joy.

## 4.0 Puzzles Solved

The theory of interpretation strategies helps answer the questions posed earlier. We can explain when loose ends will be left and when inferences will be made as being dependent upon which inference path the individual chooses. This explanation also suffices for predicting when inferences will be supplanted. Another problem that this model addresses is determining which of two equally plausible and parsimonious interpretations will be selected. Along these lines, Schank, et. al. explain such misunderstanding in verbal communication by "...maintaining that deriving a point is a part of processing, specifically related to the choice of an 'inference path'. Understanders choose to process idiosyncratically" (Schank, et.al., 1982, p. 263). This explanation of deriving a point agrees with our theory of inference paths. However, rather than believing understanders' processing to be idiosyncratic, this model predicts that individuals will tend to follow a single strategy consistently, rather than arbitrarily switching from path to path.

Another puzzle is presented in Rumelhart's (1981) work. His subjects had stories presented either a word at a time, a line at a time, or all at once. The subjects' inferences were collected either at the end of a line or at the end of the story. Rumelhart compared final interpretations of subjects who read the whole story, and subjects who read the story a line at a time, and wrote:

The results showed that subjects who interpreted a line at a time nearly always generated the same interpretations as those who gave us an after-the-fact interpretation. The only discernable difference was that those who gave an interpretation only at the end showed somewhat more variability in their interpretations (p. 27).

Rumelhart's own explanation of this phenomenon attempted to write it off as 'carelessness' on the part of the subjects:

It appears that this results from more careless reading on the part of the subjects offering an interpretation only at the end (p. 27).

However, viewed in terms of the inference strategies of Judgmental Inference theory, it is possible to interpret Rumelhart's data as further evidence for the hypothesized strategy paths. When text is presented to subjects a line at a time, with inferences about each line required, subjects are forced to act like Perseverers. Even if no inferences were elicited after each line, other demand characteristics of the task virtually force the subject to interpret the text in a particular manner. For example, if the text is presented a single line at a time, visual cues which would allow the subjects to recognize that there is more text available which might guide inferences would be lost. When subjects have all the text presented at once, they are free to interpret text using their usual strategy paths. Thus, the greater variability of

interpretation is an artifact of these different strategy paths, not a result of careless reading.

## 4.1 Experimental Validation of Hypothesis

Experiments are being conducted to discover whether the inference strategies exist, what characteristics should be ascribed to them, and whether individuals tend to use only one of the inference paths. In general, our experimental methodology is similar to that which Seifert, et. al. (1982) used in their experiments on pragmatic inferences. Our experiments, like theirs, utilize the methods of false recognition of material not found in the text, timing subjects' reading speed, and inquiring about the subjects' inferences only after a full text is read. The main difference was that Seifert, et. al., used texts of 17 lines each. They reasoned that it was possible that when readers were presented with only two lines of text, their inference strategies may be different than when reading a longer text; i.e., they may see it as only a story fragment, whereas a longer text looks like a full story.

The stories used in our experiment are not as long. However, we have controlled for the possible 'isolation effect' of short texts. The control stories used in this experiment varied in length: some of the stories were as short as the diagnostic stories, while some were several lines longer. If there is a difference in processing found between the long and short control stories, then it is likely that the 'isolation effect' is taking place in subject's analyses of the experimental texts. These techniques should make our experimental results externally valid.

## 4.2 Materials

Ten story sets, each consisting of one story and between six and nine questions, are presented to each subject as a single trial. Each story described a fairly stereotypical situation found in literature and the media. There were two kinds of questions to be answered. The first type of question required the subject to provide either information given in the story or an inference about the situation described in the story. The second type of question required the subject to make a truth judgment about the information in the question, which was either about information from the story or about inferences that could be made about events in the story.

Each trial had five control and five experimental story sets. Control stories were written to virtually force one shape of interpretation about the story situation; they were worded so that inferences would be made at the same points by everyone. Experimental, or *diagnostic*, stories were worded so that different shapes of interpretations are possible, and so that inferences need not be made at the same points by all readers, depending upon how the reader processes the story. Usually, the diagnostic stories allowed two nearly opposite shapes of interpretation. Also, the sentence order of experimental stories would permit rearrangement with the same shapes of interpretation possible, whereas the control stories' sentences could not be rearranged without destroying their sense.

Four versions of each diagnostic story were used. Some versions were permutations of sentence order, as explained above. Others had additional information which forced the shape of interpretation, but still allowed inferences to be made at different places in the text by readers using different. Because some interpretation shapes may tend to be more common than others, or interpretation shapes may be applied in a particular order, the different versions of the diagnostic stories had default inferences corresponding to several of the interpretation shapes. Thus, differences in subjects' interpretations could be accounted for by different processing methods, rather than particular interpretation shape biases. Only one version of each story appeared in a trial, so four trials with the different versions were constructed.

In addition to the different versions of the story, there were three *methods of presentation*. A story was either presented in its entirety or one sentence at a time, to test Rumelhart's results, discussed above. Asterisks (\*) were in the text at inference points and at the end of every sentence. However, when the stories were presented a sentence at a time, the method of presentation was either with asterisks at inference points and at the ends of sentences, or only at the ends of sentences. Thus, there were three methods of presentation possible of the four versions of the story sets, for a total of twelve trials.

### 4.3 Procedure

Subjects were run individually. The subjects read instructions from an Apple II microcomputer, which informed them that they would take a reading comprehension test. The subjects were told to read the stories for comprehension rather than speed. They were to press the return key as soon as they read past an asterisk, either in the text or after a question. Subjects were instructed that all questions were to be answered; responses such as "I don't know", or "the story didn't say" were prohibited. Subjects were encouraged to answer with their best guess if they weren't sure of the correct response. It was suggested that the subject think of the stories as "situations", rather than respond with the actual text of the story.

No task intervened between a story and the questions. Reading time was recorded at each asterisk, which were placed at inference points, ends of sentences, and ends of questions. One question in each story set could be (randomly) re-presented, following the statement "That's a good answer, but there is a better one. Can you think of it?". Both answers to these questions would be recorded, as well as the order in which the questions were presented. A subject could be re-questioned from zero to five times, the number chosen randomly. Each subject was given as much time as necessary to complete the trial.

### 4.4 Results

Not all the data for this experiment have been collected yet. However, preliminary results indicate that the Recency and Perseverer strategy paths do exist. The theorized characteristics of both groups are also seen to be supported. In particular, Recencies leave loose ends when no goal or plan is stated, and make inferences consistent with the default inference of the latest text, as evidenced so far by reading times and question-answering data. Furthermore, Perseverers make initial inferences about goals and plans, and cling to the initial inferences whenever feasible. It also appears that each individual tends to favor a single inference strategy.

When text is presented to subjects a single line at a time, the final interpretations tend to be more uniform than when the text is presented all at once. Though it is still unknown if this effect is a significant one, this confirms Rumelhart's (1981) findings, and is consistent with the theory that without cues about text length, subjects are forced to act as Perseverers. This is evidence that strategy paths can be chosen by readers, and hence that different choices of inference paths are *not* due simply to individual differences, but are the result of distinct strategies.

## 5.0 Conclusion: Future Work

The experiments described here were only designed to confirm that the strategy paths exist by demonstrating that readers using the different strategies have different reading and understanding behavior, most notably, completely different interpretations of particular texts. However, these experiments did not carefully explore the underlying rules each strategy has which govern inference decisions. Other experiments are currently being designed to test hypotheses about the nature of these rules.

We have constructed a prototype for a computer program which models the Perseverer and Recency inference strategies, called STRATEGIST, described in Granger, Eiselt, & Holbrook (1983). STRATEGIST was based on the data we have collected from the experiments discussed in this paper. We intend to extend the STRATEGIST model and use it as a test-bed for hypotheses about strategy-driven inference rules, as well as exploring the role of the inference strategies with texts of different genres.

## Acknowledgments

We acknowledge Dr. Art Graesser of California State University at Fullerton, and Dr. Ed Matthei of University of California at Irvine for many useful theoretical discussions and assistance in designing the experiments reported here. Amnon Meyers wrote the programs used to run the experiments.

## References

- Anderson, J.R. & Bower, G.H. *Human associative memory*. Washington, D.C.: Winston and Sons, 1973.
- Black, J.B. & Bern, H. Causal coherence and memory for events in narratives. *Journal of Verbal Learning and Verbal Behavior*, 1981, 20, 267-275.
- Bower, G.H., Black, J.B. & Turner, T.J. Scripts in memory for text. *Cognitive Psychology*, 1979, 1, 177-220.
- de Villiers, P.A. Imagery and theme in recall of connected discourse. *Journal of Experimental Psychology*, 1974, 103, 263-268.
- Granger, R.H. *Adaptive Understanding: Correcting Erroneous Inferences* (Report No. 171). New Haven, Conn: Department of Computer Science, Yale University, 1980. (a)
- Granger, R.H. When Expectation Fails: Towards a self-correcting inference system. *Proceedings of the First National Conference on Artificial Intelligence*, Stanford, California, 1980. (b)
- Granger, R.H. Directing and Re-Directing Inference Pursuit: Extra-Textual Influences on Text Interpretation. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence* (IJCAI), Vancouver, British Columbia, 1981.
- Granger, R.H. Judgmental Inference: A theory of Inferential Decision-Making During Understanding. *Proceedings of the Fourth Annual Conference of the Cognitive Science Society*, Ann Arbor, Michigan, 1982.
- Granger, R.H., Eiselt, K., & Holbrook, J.H. *STRATEGIST: A program that models content-driven and strategy driven behavior* (Report No. 198). Irvine, CA: Artificial Intelligence Project, University of California at Irvine, 1983.
- Haviland, S.E. & Clark, H.H. What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 515-521.
- Rumelhart, D.E. *Understanding understanding* (Report No. 100). San Diego, Calif: Center for Human Information Processing, University of California at San Diego, 1981.
- Schank, R.C. & Abelson, R.P. *Scripts, Plans, Goals, and Understanding*. Hillsdale, N.J.: Erlbaum, 1977.
- Schank, R.C., Collins, G.C., Davis, E., Johnson, P.N., Lytinen, S., & Reiser, B.J. What's the Point? *Cognitive Science*, 1982, 6, 255-275.
- Seifert, C.M., Robertson, S.P. & Black, J.B. On-Line processing of pragmatic inferences (Report No. 15). New Haven, Conn: Cognitive Science Program, Yale University, 1982.





# STRATIFICATION IN STORY

Valentina Zavarin  
University of California San Francisco

Since the 1960's studies of structures (structuralism in art) and studies of systems underlie language behavior (semiotics) have led to increased effort to formalize theoretical questions of story understanding. The decisive push in this direction is due 1) to the work of folklorists and anthropologists, particularly that of Levi-Strauss; and 2) to the discovery of Vladimir Propp and other Russian theoreticians of the 1920's (Bachtin, Shklovsky, Nikiforov, Eikhenbaum, and others). A number of studies of folklore and literary works appeared under the labels of narratology, grammar of stories, narrativics, structural analysis of narrative (Dundes, 1962, Greimas, 1966, Nathhorst, 1969, Hendricks, 1972, Petofi & Rieser, 1973, Prince, 1973, 1974, Van Dijk, 1975). Numerous studies discussed text structure and text composition (Lotman, 1977, Uspensky, 1973), typologies of formulaic and non-formulaic (nonce) texts (Rastier, 1971, Syrkin, 1975, Permyakov, 1979), the status of the narrators in the story (Booth, 1967, Dalezel, 1967, Pelc, 1971 and others). Fewer studies dealt with the aspect of the addressee or the reader (Todorov, 1970, Eco, 1979).

In search of a story understanding model, recent literature in Artificial Intelligence has raised a number of interesting questions. For example, what are the stories and what are non-stories (Black and Wilensky, 1979); what are the subordination principles between various elements of the story (goals and sub-goals), how do you tell if elements go together and what are the relations between these elements (Rummelhart, 1980); what are the elements in the story which are "candidates-for-deletion", when one models the story; is the aristotelian division of story into setting, characters, action and events a helpful segregation for present day analysis and what is the unit of analysis referred to as "event" (Black and Wilensky, 1979); how can we account for the embedding phenomenon in story or discourse; is there a primordial story structure concept (or a traditional one) which allows expectation-driven processing of stories (Mandler & Johnson, 1980); will a layered analysis of texts/stories be productive, specifically if we isolate in our analysis the level of narrative sequences from the level of the world of represented objects (Hobbs & Agar, in preparation).

I propose to outline several topics in story analysis which have received special attention in semiotics and in one way or another intersect with questions raised in artificial intelligence and cognitive science literature.

I will consider some differentiations which can be made between types of texts in grammar building. Specifically I will consider the case of differentiating texts according to whether they require an unequivocal single interpretation or not. I will then consider "within text" differentiation of strata which also may affect grammar building, as is the case with the grammars capturing events in the world of the story, as opposed to grammars capturing events of the narrative sequences (see for example stories with multiple flashbacks).

## Two Types of Texts: Type A and Type B

One of the principal typological distinctions between texts is that of the type "A" (factual text) which calls for a single interpretation and the type "B" text ("mythic" text) which allows for ambiguity and polysemantic interpretation. Type A text may also be referred to in the literature as scientific or "practical" text while type B as artistic text (see Greimas, 1968, Rastier, 1971, Syrkin, 1975, Zavarin & Coote, 1979).

## Content and Expression

In regards to the question raised by Black and Wilensky (1979) as to the kinds of knowledge which are needed to understand story content, one of the first steps is to discriminate the content of the story from whatever else there is in the story. One possibility is to start with the dichotomy of content of story vs. expression.

An important distinction between the two types of texts A and B is that in one type there is an obligatory regulation of the level of content while organization of the level of expression is optional and vice versa (Syrkin, 1975). A factual or scientific text (type A) strives towards an unequivocal organization of the plane of content and towards avoidance of contradictions, while regulation of the expression plane is not required. In a type B text (folklore or artistic text) equivocal interpretation or polysemanticism is desirable since it allows different societies, cultures and different periods to interpret a text in various ways. For example, there seems to be an infinite possibility to come up with a new interpretation of a classic artistic text (see multiple interpretations of Shakespeare's plays; see also controversial interpretations of classic works by the recent French critic Barthes, 1964). Type B text (artistic, mythic text) however, requires strict regulation of the expressive plane. Here the system of expressive means may be represented by elements of style, genre, artistic school, and the conventionalities attached to them. Different expressive systems are also superimposed by the language, its multiple cultural expressions and dialects. Artistically accomplished texts have been traditionally viewed as texts in which the artist "struck out the unnecessary", ergo the tabu of adding, or deleting something on the expression level of the artistic text. A type A (scientific) text may or may not be stylistically accomplished and may or may not have metaphorical or figurative expressions as its elements. The only obligatory feature is that this type of text must yield only one and not multiple interpretations. It should be noted that type B (mythic, artistic) text may be clear of metaphoric or other figurative devices and only be figurative (metaphoric) as a whole, yielding multiple interpretations applicable to many real life situations.\*

\*Type A texts are as well represented in folklore as are type B texts. Compare:

### Type A Texts:

- 1.--Ishodo gave Indai a pot in which poison had been prepared. Indai asks: "If I cook food in it will those who eat the food die?" Ishoko responded: "Take the pot, put hot coals in it, and put it on fire. When the coals are burned up, you can cook food in the pot." (Bushmen text)
- 2.--Rye says: "Sow me in ashes and in time." Oats says: "Stamp me in dirt and I will be the king."
3. It should be noted here that Black & Wilensky's (1979) text about how to catch a fish is a perfectly justifiable example of a story text. Procedural texts such as how to catch an animal, how to make clothes, or how to prepare food, etc. are abundant in folklore. The differences between Black and Wilensky text and a folklore one is that theirs is a nonce, type A text as opposed to a formulaic type A text found in folklore.

### Type B Text:

The Perishing of an Eagle. An eagle was flying in heaven, and shot down by an arrow, he was astonished. Who did that? He looked at the arrow and saw his own feather, he then said: "Woe is me! I am the cause of my own destruction."

Examples of contaminations of various types of texts can be exemplified as follows: the detective story demands a necessary regulation of content as all type A text; but

\*The work for this paper was facilitated by the National Institute of Mental Health Grant 31360.

### Within Text Differentiation of Strata

In proposing a story grammar Rummelhart (1980 and previous work) explains that at its basis is a theory of summarization. According to Rummelhart (1980) the important question is to determine the relevant portions of a story for summarizing. But we can always ask the question: relevant for what kind of summarizing and relevant in what way. If we consider that coherence may occur on a number of strata it is necessary to state which stratum one is subjecting to analysis and summarization. We can segregate

- 1) the stratum of sound;
- 2) the stratum of the world of represented objects-- here we could study separately within the world of represented objects, represented space, time, or alternations of points of view, etc.; next we can investigate and summarize
- 3) narrative sequences; and
- 4) the stratum of meaning units and relations between the actants.

Let us exemplify some of the levels by using the selection of the "Margie" story by Rummelhart (1980).<sup>\*</sup> I will not discuss the stratum of sound and thus start with Stratum Two.

#### The Stratum of the World of Represented Objects: The Diachronic Model (Stratum 2)

In Rummelhart's analysis, the story summary has the following schema: something happens to the protagonist, the happening triggers a goal, the protagonist gets involved in problem solving activity. The stratum of the world of represented objects has been given special attention by Propp (1968) and Greimas (1966) under the title of functional model. The Propp-Greimas model gives us an elaborate enumeration of story events or functions which may or may not be explicitly mentioned by the author. Propp's scheme can be summarized as follows: first, details about the birth of the protagonist; an initial situation which is unstable; a state of affairs with a certain contractual relation between the protagonist and the environment which is unsatisfactory; the breaking of the contract which leads to an initial event triggering a goal. After the initial event a number of changes occur which lead to a goal or to the final state. A new equilibrium is then established. A number of helping agents or antagonists modify the action along the way.

The particular "changes" have been sketched out by Propp-Greimas and Schank and Abelson in similar terms. The functional model can be compared to the model of "plans" in the Schank & Abelson tradition (Schank & Abelson, 1975, Abelson, 1975). Propp-Greimas functions (hereafter referred to as P/G) and "deltacts" in Schank & Abelson (1975) (hereafter referred to as Sch&A) may be seen in parallel. The 2 models assign special importance to:

(cont.)

it may also have elements of type B. Ancient poems of Parmenides and Lucretius and didactic writings in India and Persia are examples of contaminated types. Aphoristic literature requires regulation of both planes as exemplified by scientific works of Hippocrates, Leonardo da Vinci, and some works by Tolstoi. In humorous texts there is a strict regulation of connections, both formal and semantic, between both levels.

#### <sup>\*</sup> The Margie Story Version I.

Margie was holding tightly to the string of her beautiful new balloon. Suddenly a gust of wind caught it. The wind carried it into a tree. The balloon hit a branch and burst. Margie cried and cried.

"change in obligation to do something for somebody"-Sch&A

"change in the control of an object"-Sch&A

"change in what an actor knows"-Sch&A

"change in some quality of an object"-Sch&A (first version of 1975 paper)

"breaking of a contract"-P/G

"the communication of an object-ultimate good (on the parameter "to have")-P/G

"communication of a message" (or communication along the parameter of "to know")-P/G

"communication of quality" or enablement conditions (along the parameter "to be able")-P/G

There is also in addition to the previous functions the realization of

"change in the proximity relations of objects and actors"-Sch&A

"translocation in space"-P/G

It should be noted here that in order to perform a functional analysis on any story (analysis of the stratum of represented objects) events have to be first reconstructed in chronological order.

#### Expansion and Condensation

If we look at a skeleton of a story such as The Margie Story Version I, we know that an author has various options of expanding the story in its various parts. He may use stylistic distancing such that events in the story may be told by a special 1st person narrator. He may follow a complex chronology in what he will tell first about the world of represented objects and what will be left for last. The flow of narration—Redezeit, according to Laemmert (1967), has its own logic. There may be numerous flashbacks and embedding of narrator's testimonies about the events in the story. The author may also leave certain parts unsaid creating missing links in the story for a special effect. The narrator may take the role of an omniscient retrospective testifier about the action or there may be many testimonies by people of various intelligence and insights who will interpret and misinterpret events. In this respect a story is not different from a set of judicial protocols.

#### The Stratum of Narration Time (Stratum 3)

First person narratives may add a superstructure upon the events as they happen in the world of represented objects. This level can be exemplified by Rummelhart's summarization of the Version II Margie Story. Although Rummelhart himself believed that summaries of events which are not in chronological order lose the quality of story, we can perfectly well project Rummelhart's summary II into a story told in the first person:

Margie Story Version II (Rummelhart, 1975, with comments by the present author<sup>\*</sup>)

1. "Margie cried and cried" (said the author about a scene he observed);
2. "The balloon hit a branch and burst" (explained an

<sup>\*</sup> We are proposing the comments with reservations always to be taken when we deal not with a real story or communication but with something invented to illustrate a point as did Rummelhart. Thus no more weight should be attached to this analysis than what is granted to the original example.



- observer of the scene);
3. "The wind carried it into a tree" (said another observer);
  4. "Suddenly a gust of wind caught it" (said another observer);
  5. "Margie was holding tightly to the string of her beautiful new balloon" (said the father of Margie, and everything became clear to the author).

In real stories an author may choose to present events in chronological order or narrate them in a different order. "Margie Story I", the chronological story would normally be told in the third person by a pseudo-objective omniscient author whose personae will be hidden. "Margie Story II" would normally be told in the first person singular or plural by a narrator who is part of the world described in the story. It is a much more complex story as it incorporates Margie Story I and superimposes another level on it. The special effects created by embedding of narrators' voices should also be considered.

#### Embedding

Embedding of events (flashbacks) and narrators' voices is a common phenomenon in artistic works. Narrators' voices have been recently referred to as "registers" in psycholinguistics and journalism. Differentiation can be made between the real author-narrator and the apparent narrator (explicitly named or not). Characters in a story may become second level apparent narrators. Third level apparent narrators' testimonies may be embedded in the latter's speeches, etc. A story often consists of cascaded quotations where the real author and various apparent narrators and characters narrate about the objects in the represented world (see comment No. 1)

#### The Stratum of Meaning Units: The Paradigmatic Model (Stratum 4a)

Any story can be viewed as the author's solving of some problem about the world. In the case of the Margie Story we can infer that the elementary notions are those of "possession of beautiful objects" and "noxious acts of nature." The expanded story would have to precise the final design. The paradigmatic model of signification will then specify the deep meaning of the relation between the two terms (Greimas & Courtés, 1979).

#### Relations Between Invested Roles: The synchronic Model (Stratum 4b)

Various roles are embodied by different characters. According to Propp-Greimas "actants" are invested roles held by characters or objects. The possible relations are: one, of a teleological order--relations between the protagonist and the object of his goal; two, relation of an etiological order may be seen between the figure that sets the goal ("donor") and the "obtainer" of the new established order (society, for example); three, antagonists and helpers are the modifying forces in the action (Propp, 1968, Greimas, 1966). These are the basic roles which the author can distribute amongst multiple characters and significant objects which change the action.

#### Generation of Narratives

The question we can ask now is what would a full blown text of the Margie Story look like? What are the possible versions which can be generated from the given schema.

In discussing the generative aspect of narrative grammars Greimas (1971) raises the question how to account for various intermediary processes which lead from the deepest narrative structure level to the surface struc-

ture. Under the manifest level of the narrative (which only hides the signification), we find narrative structures consisting of syntagmatic chains of functions (Propp's functions & deltaxes). Under the units of the narrative syntax (defined by Propp) we find the deep meaning of the myth-like signification of the narrative (paradigmatic signification defined by Levi-Strauss).

Between the deep structure level and the linguistic surface structures are various sub-levels: (1) the level of temporalization and spatialization, (2) linguistic expansion and condensation which accounts for elision and presupposition of some narrative blocks as well as the expansion and multiple re-enactment of other narrative blocks. Finally we have (3) the stylistic distancing level which superimposes metaphors, metonymy, and other figurative devices masking the content. The realization of each of these levels in their various forms is available to the author. A simple schema such as the Margie story can become a newspaper factual account, a poem, a myth, a tragedy or a comedy.

Particularly in artistic works, the level which the author selects for presentation of the deep narrative structure, depends on the intended effect upon the reader which the author is trying to achieve. (See comment No. 2)

In conclusion, one should consider that in present day discussions related to story understanding such basic questions were asked as "how comprehension might occur" (Rummelhart, 1980). The study of how to understand and structure stories has a long history if we consider the fact that Aristotle's treatise on the topic is 2,000 years old. On the one extreme we have available a problem solving model with goals and sub-goals (either for the hero or for the author). At the other extreme we have the coherence between images and scenes and in poetry coherence of sound, which leads to a different understanding of a text (Bergson, Proust). We should thus consider that a story grammar has to specify what kind of understanding one expects from an analysis or more specifically, what stratum one intends to analyze.

#### *Comment No. 1*

Pelc describes different possibilities of embedding narrators' voices in stories:

To make a distinction between the real narrator, i.e., the author of the work in question, and the character who in the text of that work "utters" the narrative monologue, let the latter be termed the apparent narrator. That apparent narrator is always a person about whom the real narrator (the author of the work) narrates. But it often happens that the real narrator narrates about the apparent narrator without mentioning him explicitly. In lyric poetry and in novels written in the form of memoirs, he uses for that purpose verb forms and personal pronouns in the first person; in epic literature he often pretends that the apparent narrator is absent; in invocatory lyric poetry he implies the existence of the apparent narrator by using verb forms and personal pronouns in the second person; and in the drama he specifies him explicitly by naming the dramatic personae...As is known, in literary works, especially in novels and stories, it happens very often that the characters narrate about something in their dialogues and/or monologues. When this happens, they become second-level apparent narrators. And if in their narratives there is a person who in turn narrates himself, we have to do with third-level apparent narrators. None of them, however, except for the author of the work in question himself, is the real narrator...When analyzing the semantic structure of narratives we have to bear in mind that we have to do with cascaded quotations.

Pelc, 1971

# A Special Case: The Reader's Model

An important distinction has been made by Mandler and Johnson (1980) about story understanding if the "reader's model" is to be considered. Does a story have a unique structure or is there a conventional structure which should be considered when we analyze the process of reading, asked the authors.

Mandler and Johnson postulated a kind of primordial story structure which allows an expectation-driven processing for the reader/listener. Knowledge of conventional structures are particularly important and *sine qua non* in a reading of artistic works. Lotman (1977) sees the model of the reader as follows:

The perception of the artistic text is always a struggle between the reader and the author...having perceived a certain part of the text, the reader

constructs the rest of the structure in his mind. The author's next "move" may confirm this conjecture ...or it may disprove this guess and demand a restructuring of the model. Each subsequent move of the author again brings about these two possibilities. And so it goes, until the moment when the author having "vanquished" the previous artistic experience, the aesthetic norms and prejudices of the reader, compels him to accept his model of the world, his view of the structure of reality. This moment of acceptance is the "closure" of the literary work; it may occur, in fact, before the end of the text, if the author uses a familiar model, the nature of which is accessible to the reader from the beginning of the work.

## References

- Abelson, R.P. Concepts for Representing Mundane Reality in Plans. In D.G. Bobrow & A. Collins (Eds.), *Representation and Understanding*. New York: Academic Press, 1975.
- Bachelard, Gaston. *The Psychoanalysis of Fire*. Boston: Beacon Press, 1964.
- Bakhtin, M. Discourse Typology in Prose. In L. Matejka and K. Pomorska (Eds.) *Readings in Russian Poetics: Formalist and Structuralist Views*. Cambridge: MIT Press, 1971.
- Barthes, Roland. *S/Z*. Paris: Seuil, 1970.
- Barthes, Roland. *On Racine*. Hill & Wang, 1964.
- Black, J.B. & R. Wilensky. An Evaluation of Story Grammars. *Cognitive Science*, 3, 1979.
- Booth, W. *The Rhetoric of Fiction*. Chicago: University of Chicago Press, 1967.
- Calame, C. & Geninasca, C. Les discours folklorique et mythique. In: J.-C. Coquet (Ed.) *Sémiotique*. L'école de Paris. Paris: Hachette, 1982.
- Dijk, T.A. van, Some Aspects of Text Grammars: A Study in Theoretical Linguistics and Poetics, The Hague: Mouton, 1973.
- Doležel, L. The Typology of the Narrator: Point of View in Fiction. In To Honor Roman Jakobson: Essays on the Occasion of his Seventieth Birthday, I. The Hague: Mouton, 1967.
- Dundes, Alan. From Etic to Emic Units in the Structural Study of Folk-tales. *Journal of American Folklore*, LXXV, 1962.
- Eco, Umberto. *The Role of the Reader*. Bloomington: Indiana University Press, 1979.
- Goffman, Erving. *Frame Analysis*. New York: Harper & Row, 1974.
- Greimas, A.J. *Sémiotique Structurale*. Paris: Larousse, 1966.
- Greimas, A.J. Narrative Grammar: Units and Levels. *Modern Language Notes*, vol. 86, Dec. 1971.
- Greimas, A.J. & Rastier, F. The Interaction of Semiotic Constraints. *Yale French Studies*. 1968.
- Greimas, A.J. & Courtés, J. *Sémiotique: Dictionnaire raisonné de la théorie du langage*. Paris: Hachette, 1979.
- Hendricks, W.O. The Structural Study of Narration: Sample Analysis. *Poetics*, 1972, 3.
- Hobbs, J.R. & Agar, M. Text Plans and World Plans in Natural Discourse (in preparation).
- Ingarden, Roman. *The Literary Work of Art*. Evanston: Northwestern Univ. Press, 1973.
- Laemmert, Eberhard. *Bauformen des Erzählens*. Stuttgart, 1967.
- Levi-Strauss, C. & Jakobson, R. Charles Baudelaire's "Les Chats." In Michael Lane (Ed.) *Structuralism: A Reader*. London: Cape, 1970.
- Lotman, JU. *The Structure of the Artistic Text*. Translated by Ronald Vroon & Gail Vroon. Ann Arbor: University of Michigan, 1977 (Translation of Lotman, 1969).
- Mandler, J.M. & Johnson, N.S. On Throwing Out the Baby with the Bathwater: A Reply to Black and Wilensky's Evaluation of Story Grammars. *Cognitive Science*, 4, 1980.
- Nathhorst, Bertel. *Formal or Structural Studies of Traditional Tales*. Stockholm Studies in Comparative Religion, No. 9. Stockholm: Kungl. Boktryckeriet P.A. Norstedt, 1969.
- Pelc, J. On the Concept of Narration. *Semiotica*, 3, 1971.
- Permyakov, G.L. *From Proverb to Folk-Tale*. Moscow: Nauka, 1979. (Translation of Permyakov, 1970).
- Petofi, J.S. & H. Rieser (Eds.) *Studies in Text Grammar*. Dordrecht, 1973.
- Prince, G. *A Grammar of Stories*. The Hague: Mouton, 1973.
- Prince, G. Narrative Signs and Tangents: Review of Claude Chabrol (Ed.) *Sémiotique narrative et textuelle*. *Diacritics*, 1974.
- Propp, Vladimir. *Morphology of the Folktale*. Louis A. Wagner & Alan Dundes Trans. Austin-London: University of Texas Press, 1968. (Translation of Propp, 1928).
- Rastier, F. Situation du récit dans une typologie des discours. *L'Homme*, vol. XI, cah. 1, 1971.
- Rumelhart, D.E. Notes on a Schema for Stories. In D.G. Bobrow & A.M. Collins (Eds.) *Representation and Understanding: Studies in Cognitive Science*. New York: Academic Press, 1975.
- Rumelhart, David E. On Evaluating Story Grammars. *Cognitive Science*, 4, 1980.
- Schank, R.C. & Abelson, R.P. Scripts, Plans, and Knowledge. Proceedings of the 4th International Conference on Artificial Intelligence, 1975.
- Syrkin, A. JA. On Certain Traits of Scientific and Artistic Texts. *Semiotica*, 14:1, 1975.
- Todorov, T. *Grammaire du Décameron*. The Hague, 1970.
- Todorov, T. *Introduction à la littérature fantastique*. Paris: Seuil, 1970.
- Uspensky, B. *A Poetics of Composition: Structure of the Artistic Text and Typology of a Compositional Form*. Trans. by Valentina Zavarin and S. Wittig. Berkeley University of California Press, 1973 (Trans. of Uspensky, 1970).
- Zavarin, V. & Coote, M. *Theory of the Formulaic Text*. Urbino, Italy: University of Urbino, 1979.

# **Paper Session #6**



# MULTIPLE MODELS OF EVAPORATION PROCESSES

Allan Collins  
Dedre Gentner

Bolt Beranek and Newman Inc.  
Cambridge, Massachusetts 02238

We have been analyzing subjects' responses to eight difficult questions about evaporation processes in order to formalize the different kinds of mental models people use in reasoning about complex systems. In this analysis we have identified three different levels of mental models that subjects use to reason about evaporation: macroscopic functional models, microscopic aggregate models, and microscopic molecular models. Models at these three levels are closely interlinked: Each dependency in a functional model is supported by one or more aggregate models, and each aggregate model by one or more molecular models.

We have represented the models at the macroscopic and aggregate levels in terms of Forbus's (1982) Qualitative Process Theory as a series of qualitative proportionalities. The proportionalities form a causal chain linking one variable to another. In the macroscopic models the linked variables are summary variables (e.g., temperature, density) that characterize masses of elements as a whole. The aggregate models link the summary variables of the macroscopic models to aggregates over space or time of individual particles. The molecular models describe the interactions of the individual particles. They are represented in terms of the incremental qualitative analysis of deKleer (1977) and Forbus (1981).

## Analysis of Subjects' Protocols

We will illustrate our analysis by comparing two subjects' responses to one of the questions about evaporation together with the correct answer to the question. We will present each subject's protocol, together with a brief description of the subject's reasoning. Then we will give our representation of models at the different levels. The question was "Why do you see your breath on a cold day?" The first subject's response was:

RS: I think again this is function of the water content of your breath that you are breathing out. On a colder day it makes what would normally be an invisible gaseous expansion of your breath (whatever), it makes it more dense. The cold temperature causes the water molecules to be more dense and that in turn makes it visible relative to the surrounding gases or relative to what your breath would be on a warmer day, when you don't get that cold effect causing the water content to be more dense. . . . So I guess I will stick with that original thinking process that it is the surrounding cold air - that the cold air surrounding your expired breath causes the breath itself (which has a high water content and well I guess carbon dioxide and whatever else a human being expels when you breathe out), causes the entire gaseous matter to become more dense and as a consequence become visible relative to the surrounding air.

At the macroscopic, functional level RS's argument is that cold air cools the breath, which causes it to be more dense, which in turn causes it to be more visible. Microscopically he suggests no mechanism for the cooling process, but in this and other answers he appears to believe a "moving crowd" model of increased density (that average distance between molecules depends on their speed) which in turn reflects a "billiard ball" model at the molecular level. He also implies that visibility at the aggregate level is a ratio of visible particles to volume of space, but does not indicate any molecular model of particle visibility.

The second subject's response to this same question was:

PC: The reason is because the air that you breathe or rather the air that you should breathe out, comes from your body and is hot air. The air which surrounds your body, because it is a cold day, will be cold air. When the hot air that you breathe meets with the cold air of the atmosphere, it will tend to vaporize almost like steam from a kettle, which of course, can be seen. Thus unlike on a hot day, when there is hot air around you and the hot air that you breathe are the same temperature, roughly, you cannot see your breath because the steam will not be formed, but on a cold day because of the variation in the temperatures and the vaporization of your breath, you can see when you are breathing. This phenomenon would not occur on a hot day because of the similarity in temperature.

At the macroscopic level PC's argument is that the vaporization rate of water in your breath depends on the temperature difference between the breath and the air. In turn the amount of steam formed depends on the vaporization rate and the visibility of the breath depends on the amount of steam formed. PC here and elsewhere equates steam with water vapor. No aggregate or molecular models are explicit in PC's answer, though he implicitly believes that vapor holds together in space and that visibility depends on the ratio of visible to invisible particles.

The actual process that leads to seeing your breath on a cold day goes as follows: The cold air cools the water vapor in the breath, which leads to a high condensation rate of the water vapor, which leads in turn to a large amount of condensed water. It is this liquid water that is visible. At the aggregate level the cooling of the water vapor is a heat-exchange process, based on "billiard ball" collisions at the molecular level. Condensation is an aggregation of water molecules around a nucleus at the aggregate level, based on dipole electrical attraction at the molecular level. The amount of water in the breath depends on condensation and dispersion at the aggregate level, which depends on the billiard ball model of molecular interaction. Finally, the visibility of the condensed water depends on the ratio of visible particles to volume of space at the aggregate level, which depends on the absorption and re-emission of photons at the molecular level.

Table 1  
Multiple Models of Why You See Your Breath on a Cold Day

	<u>Macroscopic Model</u>	<u>Aggregate Model</u>	<u>Molecular Model</u>
RS	Temp (B) $\propto_Q$ Temp (A)	-	-
	Density (B) $\propto_Q$ Temp (B)	Moving Crowd Model	Billiard Ball Model
	Visibility (B) $\propto_Q$ Density (B)	Visibility Model?	Reflectance Model?
PC	Vaporization-rate (B) $\propto_Q$ Temp (B) - Temp (A)	-	-
	Amount-of (S) $\propto_Q$ Vaporization-rate (B)	Container Model?	Billiard Ball Model?
	Visibility (B) $\propto_Q$ Amount-of (S)	Visibility Model?	Reflectance Model?
CA	Temp (V) $\propto_Q$ Temp (A)	Heat Exchange Model	Billiard Ball Model
	Condensation-rate (V) $\propto_Q$ Temp (V)	Aggregation-on-nucleii Model	Dipole Attraction
	Amount-of (W) $\propto_Q$ Condensation-rate (V)	Container Model	Billiard Ball Model
	Visibility (B) $\propto_Q$ Amount-of (W)	Visibility Model	Absorption & Re-emission

A=air, B=breath, S=steam, V=water vapor, W=water



Table 1 summarizes the three answers: that of subjects RS and PC as well as the correct answer (CA). The macroscopic view in the first column specifies the qualitative proportionalities (Q-props) that form the functional models for the three answers. These Q-props are the relations used to describe a process history in Forbus's (1982) theory. They summarize the dependencies referred to in the three paragraphs above as the macroscopic models of the process. The individuals referred to in the Q-props are specified at the bottom of the table. Where a dash appears in the Aggregate Model or Molecular Model columns, it is because it is impossible to surmise what model the subject was using. A question mark indicates uncertainty whether the subject's answer was based on a particular model.

### Aggregate Models

Table 2 shows two of the aggregate models referred to in Table 1. Each attempts to define the constraints operating on the aggregations of particles interacting over time and/or space to produce the corresponding Q-prop at the macroscopic level.

Table 2

#### Aggregate Models of Evaporation Processes

<u>Moving Crowd Model of Gases</u>	<u>Visibility Model of Suspensions</u>
$\text{Density}(M) \propto_Q - \sum_{i=1}^n \sum_{j=1}^n \text{Distance}(m_i, m_j)$	$\text{Visibility}(M) \propto_Q \sum_{i=1}^n \text{Volume}(v_i) / \text{Volume}(S)$
$\text{Distance}_t(m_i, m_j) \propto_Q \text{Speed}_{t-x, t}(m_i) + \text{Speed}_{t-x, t}(m_j)$ <p style="text-align: center;">where <math>m_i</math> and <math>m_j</math> collided at <math>t-x</math></p>	$\sum_{i=1}^n \text{Volume}(v_i) \propto_Q \text{Amount-of}(V)$ <p style="text-align: center;">where <math>M</math> = suspension of particles  <math>v_i</math> = visible particle in <math>M</math>  <math>S^1</math> = space occupied by <math>M</math>  <math>V</math> = visible matter in <math>M</math></p>
$\text{Speed}(m_i) \propto_Q \text{Temperature}(M)$ <p style="text-align: center;">where <math>M</math> = unbounded mass of a gas  <math>m</math> = molecule of gas <math>M</math></p>	

RS clearly believes that density of molecules in a gaseous state depends on the temperature of the gas, the second Q-prop for RS in Table 1. From this and other answers this appears to be supported by a moving crowd model of gasses: *The faster any particle is moving in an unbounded gas, the more distance it puts between itself and other particles. We have represented this model as a set of Q-props relating entities at the macroscopic level to aggregations of particles at the aggregate level.* The first Q-prop states that the density of a gas (at the macroscopic level) is negatively proportional to the distance between each pair of molecules. The second Q-prop states that the distance between any pair of molecules that collide at some time is proportional to their speed after they collide. The third Q-prop states that the speed of any molecule is proportional to the temperature of the gas. Thus the Moving Crowd Model relates the density of a gas to its temperature in terms of the steady state behavior of aggregates of molecules.

The second model shown represents the visibility of a mass of particles. The first Q-prop in the model states that the visibility of the particles is proportional to the ratio between the volume of the visible particles aggregated together and the volume of the space in which the mass is suspended. The second Q-prop states that the volume of the individual particles is proportional to the amount of the visible stuff in the mass (i.e., the number of the visible particles). The visibility model then relates the visibility of a mass of particles to the amount of visible material in the mass.

These models exemplify how we have tried to capture understanding at the aggregate level. Each model relies on mappings between functional quantities at the macroscopic level (e.g., temperature of a gas) and aggregate quantities at the microscopic level (e.g., average speed of the particles in the gas). By mapping down to the aggregate level people can "understand" a macroscopic dependency in terms of a set of dependencies at the aggregate level.

## A Molecular Model

We can illustrate a molecular model by an expert "billiard ball" model of molecular interaction. Our analysis, shown in Table 3, is an extension to colliding balls of the incremental qualitative analysis of deKleer (1977) for rolling balls and Forbus (1980) for bouncing balls. All possible collisions of two balls of equal mass are summarized by the four cases shown and their combinations.

Table 3  
Billiard Ball Model of Molecular Interaction

Case 1

Entry

Exit

Impact angle	m <sub>1</sub>		m <sub>2</sub>	
	x	y	x	y
	-a	0	b	0
0°	b	0	-a	0
45° b=a b=0	0 -1/2 a	a 1/2 a	0 -1/2 a	-a -1/2 a
90°	-a	0	b	0

Case 3

Entry

Exit

Impact angle	m <sub>1</sub>		m <sub>2</sub>	
	x	y	x	y
	-a	0	0	b
0°	0	0	-a	b
45° b=a b=0	0 -1/2 a	a 1/2 a	-a -1/2 a	0 -1/2 a
90°	-a	b	0	0

Case 2

Entry

Exit

Impact angle	m <sub>1</sub>		m <sub>2</sub>	
	x	y	x	y
	-a	0	-b	0
0°	-b	0	-a	0
45° b=a b=0	-a -1/2 a	0 1/2 a	-a -1/2 a	0 -1/2 a
90°	-a	0	-b	0

Case 4

Entry

Exit

Impact angle	m <sub>1</sub>		m <sub>2</sub>	
	x	y	x	y
	-a	0	0	-b
0°	0	0	-a	-b
45° b=a b=0	-1/2 a -1/2 a	a 1/2 a	-1/2 a -1/2 a	a 1/2 a
90°	-a	0	0	-b

When two balls collide, the ball moving faster initially is defined to be  $m_1$ . The orientation of the X axis is defined by the direction of  $m_1$  (negative in the X direction). The result of the impact is defined for four critical directions  $m_2$  may be moving with respect to the X axis: left, right, up, down. Other possible trajectories of  $m_2$  are additive combinations of two of these cases: e.g., up and left, up and right, etc.

The point of impact on  $m_2$  is defined by an impact angle, measured from the center of  $m_2$  as the origin, one side parallel to the X axis and the other side defined by the contact point. We have defined the result of impact for each of three critical angles: 0 degrees where for Case 1 the two balls meet head on, 45 degrees where the line from the center of  $m_2$  to  $m_1$  has a slope of 45 degrees, and 90 degrees where the two balls just barely touch each other. Other possible impact angles have values intermediate between these three angles: These are the critical angles for inferring what will happen when two balls collide.

Let us explain the table in terms of what happens in Case 1. The two balls come toward each other each with a component of velocity in the x direction, but none in the y direction. Therefore, the entry velocity for  $m_1$  has  $-a$  for its x-component (minus because it is headed in a negative direction) and 0 for its y-component. Similarly, the entry velocity of  $m_2$  has  $b$  for its x-component and 0 for its y-component. If the two particles meet head on (i.e., their impact angle=0 degrees), they exchange momentum. Thus,  $m_1$  goes off to the right with velocity  $b$  and  $m_2$  goes off to the left with velocity  $-a$ .

If the two particles meet at a 45 degree angle, then there is a range of possible outcomes. The two boundary conditions ( $|b|=|a|$  and  $b=0$ ) for that range of outcomes are shown:  $b$  cannot be greater than  $a$  because  $m_1$  is arbitrarily defined as the faster moving ball. When  $b=a$ , after the balls collide  $m_1$  goes straight up with a velocity of  $a$  and  $m_2$  goes straight down with a velocity of  $a$ . When  $b=0$  (i.e.,  $m_2$  is stationary),  $m_1$  goes off at a 135 degree angle with its x and y-components of velocity each  $1/2 a$  and  $m_2$  goes off at 225 degrees with the same components of velocity. As  $b$  increases from 0 to  $a$ , the angle at which  $m_1$  goes off moves from 135 degrees to 90 degrees, its

x-component of velocity in absolute terms decreases from  $1/2 a$  to 0, and its y-component of velocity increases from  $1/2 a$  to  $a$ . Similarly, for  $m_2$  the exit angle changes from 270 degrees to 225 degrees as  $b$  increases, the x-component of velocity decreases from  $1/2 a$  to 0 and the y-component of velocity increases from  $1/2 a$  to  $a$ .

Naive models of billiard ball interaction are not this sophisticated, but they can be represented in similar terms. For example, a naive model might not assume momentum transfer in a head-on collision. A naive person might assume rather that the input speed for each particle is the same as its output speed in such a collision. Furthermore, naive people may not know what happens in some of the cases, or have only approximate bounds on what will happen. A more qualitative representation related to this one would give the output angles and whether the velocity is zero or not. This may come closer to the way people intuit particle interaction. But we think the parsing into cases, the combining of cases, and interpolating values corresponds to the way people think about particle interaction at a molecular level.

### Conclusion

When we looked in detail at people's reasoning about evaporation, we found that they reason at three distinct levels: (a) in terms of macroscopic variables like temperature, density, or volume, (b) in terms of aggregates of particles that behave in a similar way, and (c) in terms of individual particles and their interactions. We have tried to show how people's models at these different levels can be represented in terms of the Qualitative Process Theory of Forbus (1982) and the Incremental Qualitative Analysis of deKleer (deKleer, 1977). In particular we would argue that in principle each step in a macroscopic functional model is supported by one or more aggregate models, and in turn each aggregate model is supported by one or more molecular models.

Our study perhaps raises more questions than it answers. One important question is how many different kinds of models people have at each level of analysis. Our guess is that there are many such models, since they reflect knowledge that subjects learn throughout their lifetimes. The commonality between subjects will be in the levels at which such models are constructed and the internal language in which they are constructed.

### References

- de Kleer, J. Multiple representations of knowledge in a mechanics problem solver. Proceedings of the Fifth International Joint Conference on Artificial Intelligence Cambridge, Mass.: Massachusetts Institute of Technology, 1977, 299-304.
- Forbus, K. A study of qualitative and geometric knowledge in reasoning about motion (MIT-AI Technical Report 615). Cambridge, Mass.: Massachusetts Institute of Technology, 1981.
- Forbus, K. Qualitative process theory (MIT-AI Memo 664). Cambridge, Mass.: Massachusetts Institute of Technology, 1982.



A CENTRAL PROBLEM IN REPRESENTING HUMAN KNOWLEDGE  
IN ARTIFICIAL SYSTEMS: THE TRANSFORMATION OF  
INTRINSIC INTO EXTRINSIC REPRESENTATIONS

Gerhard Dirlich<sup>\*</sup>, Christian Freksa<sup>\*</sup>, Ulrich Furbach<sup>f</sup>

<sup>\*</sup> Max-Planck-Institut für Psychiatrie, Munich, West Germany

<sup>f</sup> Hochschule der Bundeswehr München, Neubiberg, West Germany

THREE PROBLEM SOLVING SITUATIONS

A well-known non-verbal intelligence test is the Standard Progressive Matrices, SPM (RAVEN 1958). In each test problem, the subject has to find the element in the solution set which completes the test pattern in a globally consistent way (Fig. 1).

There are two versions of the SPM. One version allows physical manipulation of the solution set, i.e., paper cards can be physically inserted into a slot in the test pattern (situation A). In the other version, test pattern and solution set are printed on the same page of a booklet, so that the subject can visually inspect test pattern and solution set (situation B).

Let us consider an hypothetical situation where an artificial problem solving system plays the puzzle (situation C).

In all three situations the problem is given in the physical world. In situation A, the problem can be solved by physical manipulation of the object. The subject has solved the problem as soon as he has obtained a "good figure" (c.f. ANDERSON 1980, pp. 53-54). Note that the subject does not necessarily need a representation of the problem, as such, in his mind. That is, he may manipulate each of the eight objects and only perceive the problem, post hoc. In contrast, situation B requires some mental representation of the problem. By applying cognitive procedures, the physical problem solving process can be mentally simulated. Then, the mental solution triggers an appropriate action in the physical problem world. In situation C, the problem must be represented in the domain of the artificial system.

In the present paper we will discuss the general structure of situation C in the light of cognitive representation theory (PALMER 1978). We believe that this can clarify some aspects of the representation of human knowledge in artificial systems. We suggest issues in knowledge representation which require further study for a better understanding of intelligent behavior.

## MEDIA OF REPRESENTATION

A close-up look at the structure of the problem solving procedures in situations A, B, and C reveals some important differences (Fig. 2).

They differ with respect to the structure of the problem solving procedure, especially with the types of representation involved. The problem solving procedure in situation C can be broken up into a sequence of subtasks: 1.) the creation of a symbolic representation of the physical problem (c in Fig. 2C), and 2.) its transformation into a computational representation (d in Fig. 2C). In case a symbolic representation of the problem is generated outside the symbol processing system (solid box in Fig. 2C), the task of the system is a classical symbol processing task. The preceding generation of the symbolic representation of the real-world problem must still be done by a human.

Is it possible to move the interface of the system in between real-world problem and its symbolic representation? In other words: is it possible to create symbolic representations of real-world problems automatically? In the following, we will discuss this question from the perspective of representation theory.

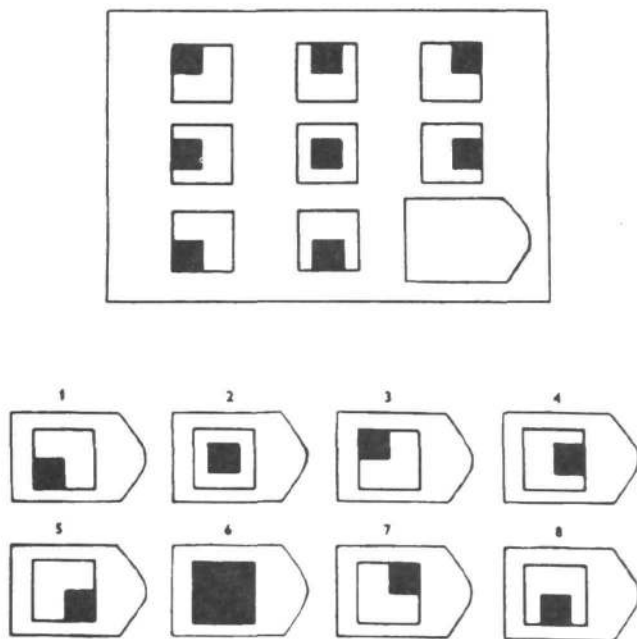


Fig. 1:  
Problem C7 from the Standard Progressive Matrices intelligence test. Upper part: test pattern, missing field at lower right corner.  
Lower part: multiple choice solution set.

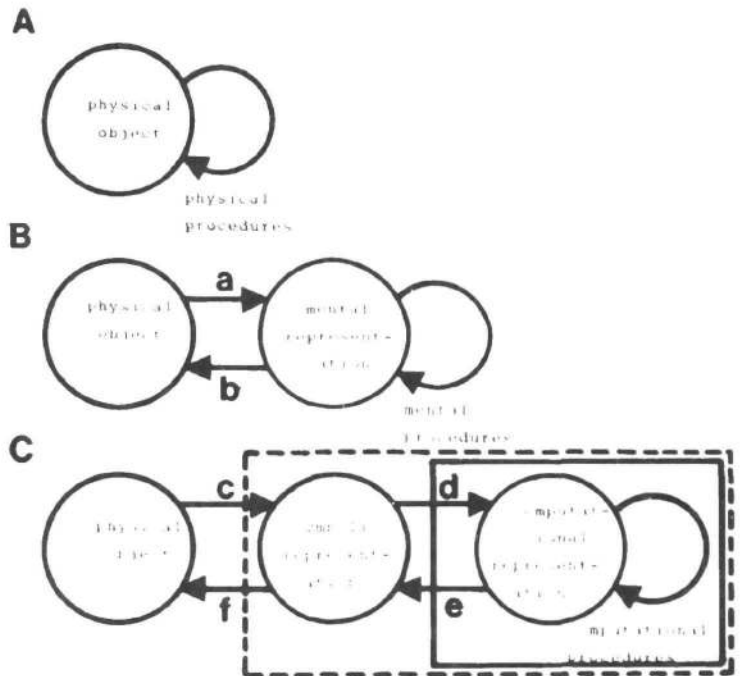


Fig. 2:  
Three problem solving situations. The problem is given in the physical world. In situation A it is solved by physical procedures. In situation B the problem and its solution is mentally simulated. In situation C, a symbolic representation of the problem is generated which then is fed into an artificial problem solving system.



## TRANSFORMATION OF INTRINSIC INTO EXTRINSIC REPRESENTATIONS

Representation theory appears to be a powerful tool for a more detailed study of the generation of symbolic representations of problems in the real world. Following Palmer (1978), a representation system mainly consists of two related, but functionally separate, worlds. In order to specify a representation system, five aspects have to be defined:

1. what the represented world is,
2. what the representing world is,
3. what aspects of the represented world are being modeled,
4. what aspects of the representing world are doing the modelling,
5. what the correspondences between the two worlds are.

The two worlds in such a representation system consist of objects that are characterized by the relations among them. The correspondence between the represented world and the representing world must preserve at least some of these relations. Two fundamentally different forms of representation should be distinguished: intrinsic and extrinsic ones. A representation is called intrinsic whenever a representing relation has the same inherent constraints as its represented relation; it is called extrinsic whenever the inherent structure of a representing relation is arbitrary and that of the represented relation is not.

Let us assume here that the creation of a symbolic representation is a two-step process: in the first step, an analog, "natural isomorphism" - type representation (SHEPARD 1975) is created. In a second step, the analog representation is transformed into a symbolic representation. A symbolic representation is a propositional representation which is sufficiently complete for solving the problem.

Palmer (1978) has argued that analog representations are intrinsic, whereas propositional representations are extrinsic. Thus, the just stated assumption allows for an interesting conclusion: the transformation of the analog representation into a propositional representation requires the generation of an extrinsic representation from an intrinsic one; this process is a crucial step in generating a symbolic representation.

The entire problem solving process is now segmented into three major subtasks. One subtask is the transformation of an analog representation into a propositional, i.e. symbolic, representation. We have just clarified that this process transforms an intrinsic representation into an extrinsic one. The other two subtasks, the generation of the initial analog representation (cf. MARR 1976) and the symbolic problem solving procedure (cf. SIMON 1978) are relatively well understood.

## ANALOG - PROPOSITIONAL DUALISM

Analog and propositional representations are regarded as the two major candidates of representations in human memory. We will briefly point out some of the impacts of the so-called analog - propositional controversy in psychology (PYLYSHYN 1973, KOSSLYN 1976) on artificial

intelligence. Speculations about the nature of representations in human memory, according to which they are either analog or propositional, can be traced back to pre-scientific times. Psychologists have come up with empirical support for either type of representation. The state of the art is probably best described as a growing belief that "analog" and "propositional" describe differing appearances of a unique underlying form of representation.

It might be appropriate to view our present understanding of representation in memory not as a controversy but rather as an analog - propositional dualism much like the wave - particle dualism of light in physics.

If we assume this point of view, some questions are shifted into the focus of interest which have not been dealt with extensively, so far: 1.) What are the conditions under which knowledge appears in analog or propositional form, respectively? This is a question aimed at cognitive psychologists; 2.) How can transformations between one form and the other be described formally? This is a question aimed at AI. There have been attempts to develop partial ad-hoc solutions within special domains (e.g. WINSTON 1975). However, to the knowledge of the authors, there have been no sufficiently general approaches on the level of representation theory.

## CONCLUSION

Let us return to the picture puzzle depicted in figure 1. What would a propositional representation of this problem look like? Clearly, many symbolic representations are possible. We outline one of them:

The test pattern consists of 9 fields. The position of each field can be propositionally described by its row ( $x$ ) and its column ( $y$ ).

The position of the black square relative to its field can be described with respect to three rows ( $x'$ ) and three columns ( $y'$ ).

Obviously, the proposition  $(x' = x) \ \& \ (y' = y)$  holds for the 8 complete fields of the test pattern.

Thus, one must conclude that it should also hold for the missing field. Therefore, pattern 5 in the solution set is correct.

The idea behind this representation can be summarized as follows: the position of the black square relative to its field is the same as the position of the field relative to the entire test pattern. Obviously, the propositional representation does not immediately pop out of the picture. On the contrary, once the solution of the puzzle is found it appears as an artificial, not easily graspable post-hoc justification of the decision.

The propositional depiction of the problem given above is based on a number of properties which are intrinsically represented in the analog representation given by figure 1. The complexity of the propositional representation stems from the difficulties we have in transforming intrinsic representations into extrinsic ones. In other words, it may

require more mental effort and more intelligence to translate the problem into a representation which is appropriate for our present symbol-based problem-solving procedures than to solve the problem in other ways. Two alternative approaches are conceivable: 1.) to design intelligent devices which can directly operate on analog representations, or 2.) to explore the problem of transforming intrinsic into extrinsic representations and to develop algorithms which can perform such transformations.

#### ACKNOWLEDGEMENTS

The issues discussed here were stimulated by fruitful interactions with Don Norman, Dave Rummelhart and their research group at the Cognitive Science Institute, University of California, San Diego, while one of the authors was a visiting scholar.

We thank Scott Campbell for useful comments and linguistic assistance.

#### REFERENCES

- ANDERSON, J.R. (1980), Cognitive psychology and its implications. San Francisco: Freeman.
- KOSSLYN, S.M. (1976), Can imagery be distinguished from other forms of representation? Evidence from studies of information retrieval times. *Memory & Cognition* 4, 291-297.
- MARR, D. (1976), Early processing of visual information. *Philosophical Transactions of the Royal Society. Series B*, 275, 483-524.
- PALMER, S.E. (1978), Fundamental aspects of cognitive representation. In: Rosch, E. & Lloyd, B.B. (eds.), *Cognition and categorization*. Hillsdale: Lawrence Erlbaum.
- PYLYSHYN, Z.W. (1973), What the mind's eye tells the mind's brain: a critique of mental imagery. *Psych. Bull.* 80, 1-24.
- RAVEN, J.C. (1958), *Standard progressive matrices*, sets A, B, C, D, and E, London: H.K. Lewis.
- SHEPARD, R.N. (1975), Form, formation, and transformation of internal representation. In: Solso, R.L. (ed.), *Information processing and cognition: The Loyola Symposium*. Hillsdale, N.J.: Lawrence Erlbaum.
- SIMON, H.A. (1978), Information-processing theory of human problem - solving. In: Estes, W.K. (ed.), *Handbook of learning and cognitive processes*. Hillsdale, N.J.: Lawrence Erlbaum.
- WINSTON, P.H. (1975), Learning structural descriptions from examples. In: Winston, P.H. (ed.), *The psychology of computer vision*. New York: McGraw-Hill.



WHAT ELSE IS WRONG WITH NON-MONOTONIC LOGICS?  
Representational and Informational Shortcomings

Jane Terry Nutter  
Department of Computer Science  
State University of New York at Buffalo  
Amherst, New York 14226

ABSTRACT

Non-monotonic logics have been used recently for a variety of A.I. purposes, including belief revision and default reasoning in question-answering and expert systems. This paper argues that by their nature, such systems discard information which has a role in human belief systems. In particular, systems which use non-monotonic reasoning lose the distinction between fully justified inferences and reasonable presumptions, in the process losing the ability to record failed expectations as such, an ability which provides a useful measure of salience for A.I. systems.

INTRODUCTION

Many A.I. systems deal less with knowledge than with beliefs which are incomplete, which change, and which frequently include generalizations which are known sometimes to fail. Consequently, it sometimes seems not only desirable but necessary to draw conclusions which are not strictly entailed by the information in the system and which further information might counterindicate. In traditional logics, if a set of premises permits a conclusion to be inferred, any larger set containing the original premises permits the same inference. Logics having this property are called monotonic. Because reasoning from beliefs seems not to share this property, researchers have devised non-monotonic logics for use in various A.I. systems.

Non-monotonic logics have generally been founded either on principles from modal logic (see e.g. [McDermott and Doyle 1980] and [McDermott 1982]) or on Zadeh's work [Zadeh 1965; Zadeh 1968] on fuzzy sets and fuzzy logic (see e.g. [Aronson, Jacobs and Minker, 1980]). A substantial technical literature now provides model theories for and proves theorems about such logics (see e.g. [Lee 1972] and [Reiter 1980]), and several A.I. systems have implemented some degree of non-monotonic reasoning, with or without some additional reasoning scheme (see e.g. [Duda, Hart, Nilsson and Sutherland 1978]). Not all responses have been favorable. Non-monotonic logics have been criticized both for technical shortcomings (e.g. [Davis 1980]) and on more general philosophical grounds, which question using logic to govern inferences which involve complex judgements of causal connections and the like (see e.g. Israel 1980]).

The criticisms I want to put forward here fall into neither of these categories. Instead they have to do with the precise role such logics play in A.I. systems when they are used to model reasoning from default generalizations. I will argue that non-monotonic logics do not and in principle cannot accurately reflect such reasoning. Obviously, this criticism depends strongly on what is to be modeled. Hence the first step is to get clear about what uses of non-monotonic logic I am criticizing. The second section discusses this issue. In the third section, I take up the particular aspects of knowledge and information based on default reasoning which I claim non-monotonic logics by their nature cannot reflect.



## THE PHENOMENA

It is not always immediately clear what a proposed system is intended to model. Fuzzy logic in particular has been associated with many different purposes: modelling inference patterns, associating probability measures with events or confidence measures with propositions (neither of which is the same as modelling inference patterns), measuring class membership ("To what extent can we call a bacteriophage an animal?"), and almost anything else which someone might want to measure using numbers between zero and one. Each of these uses raises questions of appropriateness; certainly I don't mean to take them all on here.

I am interested specifically in the use of non-monotonic logics to implement inference in systems which are intended either to model some belief space or to perform as expert systems with extended capacities that would include, for instance, formulating descriptions of things they know about (including facts they learn), explaining what makes some object or event unusual, and so on. The sorts of systems I am talking about can receive new information, store representations of it, perform inferences using new information, and report the results of all these operations in some reasonable fashion. They also perform reasoning based on generalizations in situations of incomplete information ("default reasoning"). At least in principle, they can represent any information in their domains in which a human might be interested, and can form propositions expressing that information (though of course not necessarily the same sentences that a human agent would use). By "in principle" I mean that while they may currently lack these abilities, the abilities are desirable, and it is anticipated that they could be added to the system, although it might take some research to find out exactly how.

Such systems could in principle use non-monotonic logic many different ways. This paper deals with using non-monotonic logic as the basic inferential mechanism for such systems. My claim is this: non-monotonic logics cannot support the capabilities listed above. Moreover, this inability does not stem from any technical feature of existing non-monotonic logics. Rather it results directly from non-monotonicity itself when applied to default reasoning.

## THE SHORTCOMINGS

1. Their identity. Suppose we start in out in situation  $S$ , with information (including unqualified beliefs and generalizations) supporting but not strictly entailing conclusion  $C$ . For instance, we know what birds are, and we know that Roger is a living unplucked bird. This information supports the conclusion that Roger flies, but we know this might fail. Now suppose that we learn something which contradicts our previously supported conclusion. In our example, we might learn that Roger is a kiwi. Call the new situation  $S^*$ . According to non-monotonic logic,  $S$  entails  $C$ ,  $S$  does not entail not- $C$ ,  $S^*$  entails not- $C$ , and  $S^*$  does not entail  $C$ . This set of entailment relations provides a consistent basis for stating whether or not, to the best of our current knowledge, Roger flies. But it is not a sufficient basis for all of our relevant knowledge in either  $S$  or  $S^*$ .

In situation  $S$ , a careful speaker would not say "Roger flies," but something like "It is reasonable to suppose that Roger flies." This qualification would not be placed on all conclusions about Roger: "Roger has feathers" is justified absolutely by our beliefs in  $S$ . We may be mistaken that Roger is a live unplucked bird, but if we are right about that, we must be right about his having feathers, since by the biological definition of birds, they all without exception have a genetic disposition to produce feathers.



The relationship between our knowledge and the conclusion that Roger has feathers is fundamentally different from the relationship between our knowledge and the conclusion that Roger flies. Furthermore, people are frequently interested in this kind of difference: questions like "Are you sure?" would not otherwise be so common. There is an important distinction between expectations and knowledge. But non-monotonic logic treats the relationship in both cases as entailment. Worse, it produces exactly the same conclusion in situation S as it would in situation S\*\*, where we add the information that Roger is a mature parrot, uncrippled and in normal health. But we know that Roger-the-parrot flies, whereas we only suspect that Roger-the-bird flies. We already know in situation S that this uncertainty exists. This is the first shortcoming of non-monotonic logic: it loses the distinction, present in the "real life" situation, between justified beliefs and justified assumptions. (For more on this distinction, see [Nutter 1982].)

In situation S\*, more information is lost. When we learn that Roger is a kiwi, we don't just know that he doesn't fly. We know that (a) Roger is a bird; so (b) there is reason to believe that Roger flies; but in fact (c) Roger is a kiwi, so (d) he doesn't fly after all. Taken together, (b) and (d) contain substantial information: they tell us that an expectation has failed. Non-monotonic logic forces the difference between justified belief and justified assumption into the logic, so that "There is reason to suppose that Roger flies" becomes the same proposition as "Roger flies"; thus viewed, (b) and (d) represent an out-and-out contradiction. To maintain consistency, (b) must be rejected: as stated above, S\* entails not-C ("Roger does not fly") and does not entail C ("Roger flies" — in this system, the same as "There is reason to suppose that Roger flies").

So in situation S\*, non-monotonic logics can not support inferring "There is reason to suppose C, but not C". This is information in which a human might be interested. Notice that this problem arises explicitly from non-monotonicity. The information lost in S\* is precisely what is known in S. Unless that information is lost, adding premises cannot cause conclusions to be rejected, and the logic is by definition monotonic.

Hence we have two kinds of information which non-monotonic logics fail to support. First, they lose the distinction at the propositional level between knowledge and supposition. Associating a measure instead of a truth value with propositions cannot do the work needed here, especially if the metric is taken as measuring probability or confidence: the probability that a single fair toss of an unbiased coin will land heads is 0.5; and of that fact we are completely certain. Knowledge of probabilities is itself knowledge which may be either justified by other knowledge or only suggested. Furthermore, default generalizations are not statistical [Nutter 1983]. Second, because non-monotonic logics fail to support the distinction between knowledge and supposition, they also fail to support reports of failed expectations.

2. Their importance: identifying salience. Consider the following two descriptions of birds.

(a) Felix is a bird who lives in North America. He is under four feet tall, he flies, and he travels slowly on the ground.

(b) Oscar is a bird who lives in Africa. He is over four feet tall, he can't fly, but he travels very rapidly on the ground.

Which of these birds do we know more about? Felix could be almost any North American bird, except a road runner. Oscar is an ostrich, and couldn't be anything else. Yet in both cases, I have simply given the continent they live on, their height (in vague

terms), whether they fly, and their speed on the ground. How do we come to have so much more information about Oscar than about Felix?

When generalizations hold, they don't tell us much. But when they fail, their failure conveys information. Oscar's height, flightlessness, and ground speed are all unusual; together with his African origins, they pin down his species. (If he were Australian, he'd be an emu.)

Salience has been called the key to a major natural language generation problem: what and how much to say (see e.g. [Conklin and McDonald 1982]). Most techniques for determining salience depend on either marking particular properties for a class of objects or determining differences between a pair of objects (see the above, [McCoy 1982] and [McKeown 1982]). Neither of these techniques will let a system produce paragraph (b) when asked to describe Oscar but produce only the first sentence of paragraph (a) when asked to describe Felix. However, consider the following rule: if X belongs to kind K, and members of kind K typically have property P but X does not, that is interesting and should be reported. A representation and underlying logic which distinguishes "There is reason to suppose that P" from "P" and allows deducing "Not P, and there is reason to suppose that P" will support this rule for determining salience. But the logic of such a system will be monotonic.

## CONCLUSION

Non-monotonic logics have been motivated largely by the belief that reasoning from default generalizations involves denying old conclusions on the basis of new evidence consistent with all previous premises. I have argued here that this belief arises from failing to distinguish between guarded statements of the form "There is reason to suppose that P" and statements of the form "P". Given this distinction, "There is reason to suppose that P" and "not P" do not contradict one another, so on learning the second we need not reject the first. Hence a monotonic logic providing this distinction can deal with default reasoning. Furthermore, failing to distinguish between these classes of statements causes information to be lost. In particular, two specific gaps appear: (a) the system will not know and be able to report the difference between those conclusions which its premises warrant without reservation and those conclusions which its premises only suggest, and (b) because the system will lose access to the reasonable assumptions when specific information overrides them, it will be unable to detect and state that a reasonable expectation has failed. Consequently, adopting non-monotonic logics may deny A.I. systems access to a simple and useful rule for determining saliency.

## ACKNOWLEDGEMENTS

I want particularly to thank Stuart Shapiro and the members of the SNePs research group for their many helpful comments and criticisms.

## REFERENCES

1. Aronson, A.R., Jacobs, B.E., and Minker, J. A note on fuzzy deduction. JACM v. 27 (1980) 599-603.
2. Conklin, E.J. and McDonald, D.D. Salience: the key to the selection problem in natural language generation. Procs. 20th Annual Meeting of the Association for Computational Linguistics (1982) 129-135.

*Ed Center  
necessary -  
non necessary  
distinction  
be supported*

3. Davis, M. The mathematics of non-monotonic reasoning. A.I. v. 13 (1980) 73-80.
4. Duda, R.O., Hart, P.E., Nilsson, N.J., and Sutherland, G.L. Semantic network representations in rule-based inference systems. In: Pattern-Directed Inference Systems, D.A. Waterman and F. Hayes-Roth, editors, Academic Press (New York, 1978) 203-223.
5. Israel, D.J. What's wrong with non-monotonic logic? Procs. First Annual National Conf. on Artificial Intelligence, American Association for Artificial Intelligence (1980) 99-101.
6. Lee, R.C.T. Fuzzy logic and the resolution principle. JACM v. 19 (1972) 109-119.
7. McCoy, K.F. Augmenting a database knowledge representation for natural language generation. Proc. 20th Annual Meeting of the Association for Computational Linguistics (1982) 121-128.
8. McDermott, D.V. and Doyle, J. Non-monotonic logic I. A.I. v. 13 (1980) 41-72.
9. McDermott, D. Non-monotonic logic II. JACM v. 29 (1982) 33-57.
10. McKeown, K.R. The TEXT system for natural language generation: an overview. Procs. 20th Annual Meeting of the Association for Computational Linguistics (1982) 113-120.
11. Nutter, J.T. Defaults revisited, or "Tell me if you're guessing". Procs. Fourth Annual Meeting of the Cognitive Science Society (1982) 67-69.
12. Nutter, J.T. Default reasoning in A. I. systems. Draft (1983).
13. Reiter, R. A logic for default reasoning. A.I. v. 13 (1980) 81-132.
14. Smith, E.E. and Osherson, D.N. Conceptual combination and fuzzy set theory. Proc. Fourth Annual Conference of the Cognitive Science Society (1982) 47-49.
15. Zadeh, L.A. Fuzzy sets. Inf. Control v. 8 (1965) 338-353.
16. Zadeh, L.A. Fuzzy algorithms. Inf. Control v. 12 (1968) 92-102.



Tacit Theories of Economics  
William J. Salter  
Bolt Beranek and Newman Inc.

This paper presents results from work in progress on tacit theories of macroeconomics. The research can be viewed in the context of work on mental models. Virtually all of that research deals with domains where the laws of physics can be applied, where there is a single correct answer with which subjects' data can be compared, and where the underlying representation can be thought of in terms of knowledge. In the domain of macroeconomics -- as in most domains with which people have to deal in their lives, like voting, child-rearing, personal health, predicting and understanding the behavior of other people -- there is no unitary body of truth, people have beliefs and opinions as well as degrees of knowledge, and well-informed people may differ. These are heterodox rule-relevant domains: there is agreement that they are rule-governed, but disagreement about what the rules are.

The overall, long-term goals for this research project are to develop systematic, objective ways to look at mental models or tacit theories; to increase understanding of both the structure and the explanatory utility of such systems of knowledge and belief; and to contribute to addressing the crisis in which the discipline of economics finds itself, in large part as a result of unrealistic implicit psychological assumptions.

The research is still underway. In this paper, I focus on the tacit theory of one subject, as a way to illustrate both the methodology and the nature of the claims I hope to make. Data from other subjects will be considered primarily as a way of putting the single subject in context.

Several major claims are made in this paper: a small set of economic concepts -- called the "economic core" of a subject's conceptual network -- can be seen as driving an individual's understanding of how the economy works; these cores have coherent internal structures and, in effect, constitute tacit theories of macroeconomics; across people, these cores correspond well with both folk theories and crude versions of "real" theories of economics; these cores can be viewed as the machinery for computing expectations, and thus may drive not only economic inferences but perhaps some important economic behaviors as well.

The present research has four primary areas of concern:

1. Does the tacit theory of an individual remain relatively stable over the short- to medium-term? If so, one can argue that the tacit theories may reflect stable patterns of individual differences.
2. What is the internal consistency of an individual's tacit theory? I have not yet fully developed the mathematics for looking at internal consistency. Preliminary results are encouraging, and indicate high degrees of internal consistency. This topic will not be addressed further in the present paper.
3. What is the content of tacit theories? Do they make any sense from the vantage point of economics, or do people have idiosyncratic theories?
4. Are there stable patterns of individual differences in tacit theories? If the answer is no, tacit theories become of limited use as explanatory and predictive constructs, although they may still be quite interesting as objects of psychological investigation.



There are also a number of second-order questions that could be of interest. For example, consistency across time and internal consistency could be viewed as dependent variables, and one could try to explain them based on demographic factors, political beliefs, initial theory, and the like. Similarly, if there are stable patterns of individual differences in tacit theories, one could look for variables that predict the theory an individual holds.

## METHOD

Subjects for this study were 24 residents of New Haven, Connecticut, randomly selected from the telephone book. Although no claim can be made that this sample provides for valid statistical generalizations to the population at large, the sample was quite demographically diverse. Each subject was interviewed twice: first in person for about an hour and a half; after four to eight weeks, by telephone, for 20 to 30 minutes. The heart of both interviews, and all that will be discussed here, was a series of questions of the following form: "If interest rates go up, what do you think would happen to inflation? Why?"

The typical answer was a "causoid" path containing two to four intervening links. For example, a response might be: "Well, if interest rates go up, it costs the business more to borrow, so they have to pass that cost on, so prices would rise and inflation would go up." Note that there are equally plausible alternatives: "If interest rates go up, people will have to pay more on their loans, so they wouldn't want to borrow as much to buy things. And to get that money, then, businesses would have to charge less, so inflation would go down." Most of the reasoning was causal in nature, but also included diagnostic inferences (e.g., "Well, if interest rates went up, that would have to be because the deficit was high, and a high deficit means a lot of government spending, driving prices up and making inflation worse."), statements of correlation, appeals to authority (e.g., "The Federal Reserve makes those interest rates go higher to stop inflation, so I guess it would make it go down."), and occasional confusions about where the reasoning was going or had come from. Each interview was tape-recorded, transcribed, and coded.

For each subject, for each interview, a signed, directed graph was constructed, consisting of, essentially, all of the subject's answers connected together. A typical graph contains 35 nodes, of which 15 were provided by the experimenter in questions, and about sixty links. Many concepts and links were mentioned more than once.

These graphs were transformed into matrices by setting a parameter of link strength (in the data reported below this was .7) and by assuming that link strengths were multiplicative along chains. The links between concepts were extremely simple: if a change in A made B change in the same way, it was coded as a direct link from A to B, and if a change in A made B change in the opposite direction, it was coded as an inverse link from A to B. Each entry in these matrices thus represented the directed causal coefficient from the row concept to the column concept. For example, there would be three entries in the matrix for the following fragment of protocol: "If interest rates go up, people would have to pay more on their loans, so they wouldn't borrow as much." The matrix entries would be: INTEREST-RATES to COST-OF-BORROWING, +7; COST-OF-BORROWING to CONSUMER-BORROWING, -7; and INTEREST-RATES to CONSUMER-BORROWING, -49. When there was more than one entry for a cell, a formula was used to combine cell entries such that they were asymptotic to +1 or -1.

The statistical methodology was rather complex, and no attempt will be made to explain it in detail here. Other papers will treat that in some depth. The two primary analytical approaches to the data both involved the use of principal components factor analysis on each matrix. The first approach looked at only first principal component



for each matrix. Each first principal component is a vector of length equal to the number of concepts used by the subject of interest. And each entry in that vector is a good measure of the aggregate causal importance of that concept in the subject's matrix. Thus, concepts that appear early and often in causal chains will tend to have higher loadings, as the values are called, on the first principal component. The second technique looked at the four-dimensional structure embodied in the loadings on the first four principal components. Loadings on second and subsequent principal components reflect clusters of causal importance; for example, two concepts that are both causally important, but cause different effects, would not be distinguished on the first principal component -- both would have high loadings -- but would be distinguished on subsequent components. To oversimplify slightly, then, the values on the first principal component reflect causal importance, while the values on the first four principal components reflect causal structure.

To look at the extent to which causal importances between subjects (or for the same subject, between initial interview and follow-up) were similar, the loadings of the concepts on the first principal components were correlated. The higher the correlation, the more similar the relative values of the loadings. And to look at similarity of causal structures, the first four principal components from one interview were canonically correlated with those from another interview. Canonical correlation allows the two four-dimensional structures to rotate freely in order to maximize fit with each other. Thus, the canonical correlation is a summary measure of the extent to which the two causal structures are congruent.

*Can*

## RESULTS

Both modes of analysis are relevant to addressing the first question of interest, the consistency of a subject's tacit theory over time. The most direct and constrained way of looking at consistency over time is to correlate the concept loadings from a subject's initial interview with those from the follow-up, four to eight weeks later. For the first twelve subjects, these correlations ranged from .13 to .72. The lowest correlation indicates virtually no consistency. This subject repeatedly protested during the interview that she "knew nothing about economics," an on-line introspection that is supported by the data. The next lowest correlation is .44. Seven of the remaining ten are .60 or above.

In order to put these correlations in some context -- after all, it is not immediately clear that a correlation of, say, .55 is high -- the first principal components of all 24 interviews for the first twelve subjects were correlated. Of interest here is the extent to which the correlations within-subject are higher than those between-subjects. (Note that, if there are a relatively small number of importantly different tacit theories across subjects, some between-subject correlations should be quite high.) Of the 276 correlations, only 30 were .60 or above; of these 30 correlations, seven were within-subject. Nine of the 12 within-subject correlations were in the top 15 percent of the total number of correlations. I take this to be strong evidence that tacit theories are relatively stable psychological entities.

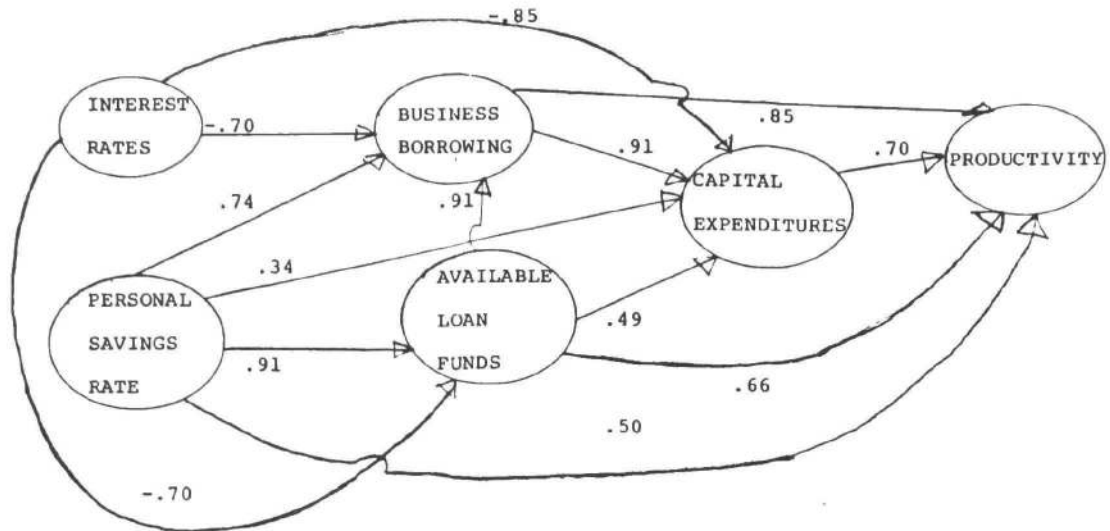
Note that concepts which a subject mentioned in one of his interviews but not in another hurt the correlation, in that they have zero loadings when not mentioned and non-zero loadings when mentioned.

The subject who will be discussed in detail was the first subject interviewed and had a correlation between his interview and follow-up of .64. A very strict criterion of semantic agreement was used to code responses. If this criterion is relaxed slightly -- collapsing the terms "investment" and "capital expenditures," which seems amply justified from the full protocols -- the correlation for this subject goes up to .76.

go down, more money becomes available for industry to borrow, borrowing increases, capital expenditures go up, and productivity climbs. Note the absence of tax rates, disposable income, the budget deficit, and consumer spending -- the key ideas in Keynesian and neo-Keynsian economics. These do enter into the cores of other subjects.

FIGURE 1:

SUBJECT 1 CORE, DIRECTED CAUSAL COEFFICIENTS FROM INITIAL INTERVIEW



The analyses of individual differences across subjects have not been completed. Based on preliminary analyses, however, it appears that there will be a small set of three to six of basic cores, with variations. These cores may well reflect stable patterns of individual differences in causal structures in the economic domain.

## DISCUSSION

Tacit theories, as operationalized in this research, appear to be legitimate and interesting objects of investigation. Subjects seem to possess a high degree of internal conceptual organization and coherence in the domain of macroeconomics, a domain in which they are non-expert and in which, in fact, they may well disagree with each other.

It may well be possible to view tacit theories, as characterized by the conceptual cores, as miniature "expectation machines," which could then serve as proxies for the endogenous measures of expectations in economic models based on rational expectations theory. This might introduce stronger psychological foundations into the models, and would also allow for a systematic treatment of individual differences, generally considered noise in economics. Such applications await further research, but offer a tantalizing promise of interdisciplinary collaboration.

Another way to look at the consistency from the initial interview to the follow-up is to look at the canonical correlation of the four-dimensional structure of the first four principal components from the interview with that from the follow-up. Canonical correlation is a way to look at the similarity of the underlying causal structures, while simple correlation of first principal components looks at similarity in the causal importance of the set of concepts. For this subject, the first canonical correlation, reduced for capitalization on chance, was .84. Thus, the two structures could be rotated into a very high degree of congruence.

Note that canonical correlation does impose a considerable degree of constraint. For example, the subject with the correlation of .13 on the first principal components had an adjusted first canonical correlation, using the first four principal components, of .24. This is statistically non-significant, and is converging evidence that, indeed, she did not have a consistent model over time.

To look at the content of tacit theories, I define the notion of a conceptual core of the tacit theory. The core contains those concepts with the highest loadings; thus, they are causally most important. In terms of the analytic procedures employed, the methods of core extraction imply that the cores are what the subject views as driving the economy. If the subject were making economic policy, and used his or her tacit theory to do it, the concepts in the core are those that the subject would attempt to influence to affect the critical endogenous variables of inflation and unemployment. The conceptual core can be computed on both the first principal components and on the results of the canonical correlations. Table 1 summarizes the concepts with loadings over .25 on the first principal components or the first canonical variate for the first subject. (The loading on the first canonical variate is a summary measure of the causal salience of each concept in the fit-maximizing rotation of the underlying four-dimensional causal structure.) In this table, investment and capital expenditures are collapsed. Note the extent to which the interview and follow-up show the same cores; note also the degree to which the results of the first principal component analysis and the canonical correlations agree.

TABLE 1:  
LOADINGS OF CONCEPTS IN SUBJECT 1's CORE

CONCEPT	FIRST PRINCIPAL COMPONENT		FIRST CANONICAL VARIATE	
	INTERVIEW	FOLLOW-UP	INTERVIEW	FOLLOW-UP
Available loan funds	.33	.32	.43	.30
Business borrowing	.38	.27	.39	.23
Capital expenditures	.33	.56	.27	.33
Interest rates	-.47	-.32	-.58	-.61
Personal savings rate	.26	.26	.34	.20
Productivity	.28	.36	.15	.08

In figure 1, the subject's core is presented as a directed graph; the numbers associated with each arrow are the directed causal coefficients between concepts in the subject's initial interview. (They are very similar for the follow-up, as the high correlations require.) The data in this figure are a subset of the data input to the extraction of principal components; thus, they are several statistical transformations prior to the correlations of the first principal components and to the canonical correlations. They provide another vantage point for looking at internal consistency and conceptual coherence, of a more qualitative sort. One can ask, of this core, does it make economic sense? It does; this core is very close to what orthodox supply-side theory (if this is not an oxymoron) specifies. As the personal savings rate goes up, and as interest rates



# Student Modeling as Strategy Learning

Pat Langley

The Robotics Institute  
Carnegie-Mellon University  
Pittsburgh, Pennsylvania 15213 USA

## 1. Introduction

One of the major prerequisites for intelligent computer aided instruction (ICAI) is the ability to construct a model of the student's knowledge. Some progress in this area has been made through carefully studying students' behavior, and deriving common "bugs" or "mal-rules" that lead to errors. Unfortunately, this approach requires that each new domain be analyzed in detail before an intelligent tutor for that domain can be undertaken. More recently, researchers have attempted to develop generative theories to explain the origin of such bugs. Such theories would help predict bugs in new domains, and lay the foundation for more general ICAI systems.

However, recent AI learning research suggests an alternate approach to student modeling. In this paper we describe SAGE, a system that learns search heuristics, and discuss its application to constructing models of students' behavior on mathematics problems. SAGE is stated as a production system, and learns by determining the conditions under which various operators should be applied. In modeling students' strategies, the system instead determines the conditions under which students apply these same operators, whether correctly or incorrectly.

## 2. Learning Search Heuristics

One of the central insights of AI is that intelligence involves *search*, and a corollary of this insight is that learning often involves the acquisition of heuristics for directing search down profitable paths. We have explored the process of learning search heuristics through SAGE, a program that begins a task with weak, general methods, and that acquires domain-specific, powerful methods as a function of experience. The system has been tested on a number of puzzle-solving tasks such as the Tower of Hanoi and Slide-Jump, as well on simple algebra problems (Langley, 1982).

SAGE is stated as an adaptive production system. In other words, its procedures are cast as condition-action rules, and it modifies its behavior by constructing new condition-action rules. The system consists of two main components. The first of these contains very general rules for assigning credit and generating new rules; this component is responsible for learning from mistakes. The second component is domain-specific, but always takes the form of rules for proposing moves through some problem space. Initially, these rules contain only the *legal* conditions for making a move. As a result, SAGE sometimes makes good moves and sometimes makes bad moves; in other words, at the outset the system must *search* for a solution. However, as experience is gained, SAGE generates more conservative versions of its move-proposing rules with additional conditions. Eventually, the program arrives at a set of heuristics that propose only useful moves and that avoid bad moves entirely; that is, SAGE learns to direct its search down desirable paths.

SAGE incorporates a number of methods for assigning credit to desirable moves and blame to undesirable ones. These include techniques for noting loops, unnecessarily long paths, dead ends, and illegal moves. However, in this paper we will focus only on the most general credit assignment heuristic: learning from complete solution paths. This method is straightforward. SAGE employs its initial move-proposing rules to search a problem space, eventually finding one or more optimal solution paths. The system then retraces its steps, marking those moves lying on the paths as good instances of the heuristics it is trying to learn. In addition, moves that lead one off the path are labeled as bad instances, since they do not lead to a solution. As good and bad instances are identified, this



information is passed to SAGE's learning mechanism, which attempts to generate heuristics for directing search.

SAGE learns through a process of discrimination. Given a bad instance for a move-proposing rule, the system retrieves the last good instance of the same rule and searches for differences between the two situations in which the rule was applied. SAGE finds all differences between these two situations, and constructs a more conservative variant of the rule for each difference that it finds, with the new conditions based on those differences. Since many differences may occur, and since some of these may be spurious, SAGE does not automatically assume that all of these variants are useful. Instead, it requires that a variant be relearned in a number of different contexts before it is allowed to direct search. This is accomplished by associating a *strength* with each rule. These strengths are initially low, but they are increased whenever a variant is relearned; once a variant's strength exceeds that of its parent, the variant is applied whenever it matches. Of course, even variants based on relevant differences may still be overly general and lead to further errors. In such cases, the discrimination process is applied recursively and still more specific rules are constructed. This process continues until SAGE can solve the problems it is given without search.

### 3. Modeling Subtraction Strategies

The application of SAGE to modeling students' strategies is straightforward. Given a student's answers to a set of problems and the legal operators for that domain, SAGE should be able to find solution paths which give the same answers as did the student. From these solution paths it should be able to determine the conditions under which that student applied the operators, using its discrimination learning algorithm. In other words, the program should be able to learn a strategy that mimics the student's behavior, and the resulting set of rules would be equivalent to a model of that student's strategy.

We have performed initial tests of SAGE's student modeling capabilities in the subtraction domain, drawing upon earlier analyses by Brown and Burton (1978), Brown and VanLehn (1980), and Young and O'Shea (1981). These researchers have identified and modeled a variety of subtraction errors, such as: (1) always borrowing, whether necessary or not; (2) borrowing when the top number is larger, but not when it is smaller; (3) subtracting the smaller number from the larger, regardless of position; and (4) pattern errors, such as  $0 - N = 0$  and  $0 - N = N$ . Our approach to modeling these errors is most similar to that of Young and O'Shea, who explained many bugs in terms of missing rules.

The application of SAGE to student modeling is best understood through an example. The system is given an initial set of rules incorporating operators for a domain. Table 1 shows the system's initial rules for subtraction.<sup>1</sup> These include operators for finding the difference between two numbers in a column, for decrementing a number by one, for adding ten to a number, and so forth. However, the rules in which these operators occur contain only the most general conditions. When the system is given a subtraction problem and a student's answer to the same problem, it must search in order to arrive at a similar answer. Once it has found a solution path to the same answer, SAGE employs its discrimination process to determine the conditions on various operators that will produce similar behavior in the future.

For example, suppose the system is given the problem  $34 - 21$ , along with the student's (correct) answer of 13. Based on the solution path leading to this answer, SAGE would give the difference-finding operator precedence over the decrement and add-ten operators (since they were

---

<sup>1</sup>We have chosen to paraphrase the rules in English for the sake of clarity. Each line in the table corresponds to a single condition or action in the actual productions. Words in italics correspond to variables in the actual rules.



never applied). In addition, the system would note that the difference-finding operator was called upon to find the difference between 4 and 1, but not the inverse difference between 1 and 4. Comparing these two situations, the discrimination routine would find two differences. In the good instance, 4 is *larger* than 1, while in the bad instance, 1 is not *larger* than 4. Similarly, in the good instance, 4 is *above* 1, while in the bad instance, 1 is not *above* 4. Accordingly, SAGE would construct two variants, one including a condition based on the *larger* relation, and the other containing a condition based on the *above* relation.

Table 2. Initial production system for subtraction.

---

**find-difference**  
 If you are processing *column*,  
     and *number1* is in *column*,  
     and *number2* is in *column*,  
 then find the difference between *number1* and *number2*,  
     and write this difference as the result for *column*.

**add-ten**  
 If you are processing *column*,  
     and *number1* is in *column*,  
     and *number2* is in *column*,  
     and *number1* is above *number2*,  
 then add ten to *number1*.

**decrement**  
 If you are processing *column*,  
     and *column2* is left of *column*,  
     and *number3* is in *column2*,  
     and *number3* is above *number4*,  
 then decrement *number3* by 1.

**shift-column**  
 If you are processing *column*,  
     and you have a result for *column*,  
     and *column2* is left of *column*,  
 then process *column2*.

---

At this point, one can infer that the student employs at least one of these conditions in deciding when to apply the difference-finding operator, but we cannot tell which of the conditions (or both) is used. However, suppose we next examine a problem in which borrowing is required, such as  $43 - 25$ . If the student gives the correct answer of 18, then SAGE would infer (after finding a solution path and discriminating) that the student's differencing rule contains both of the above conditions. However, if the student gives the answer 22 instead, then the variant including the *larger* relation would be retained in favor of the variant including the *above* relation. The resulting model would always subtract the larger number from the smaller regardless of position, and would explain the student's failure to borrow in terms of the missing *above* condition. We have described only a small part of the model-building process, since the conditions on other operators must also be determined; however, this example should give the reader an idea of the basic approach.

## 4. Directions for Future Research

Although SAGE has been tested in a number of domains as a strategy learning system, our application of the program to student modeling is still in its initial stages. The most obvious priority is to test the system more fully in the subtraction domain. After our analysis of subtraction errors has progressed, we plan to test SAGE in the domains of algebra and symbolic integration. Like subtraction, these areas are mainly procedural, but they are sufficiently different to provide an interesting test of the system's generality. Finally, we hope to provide SAGE with the ability to generate diagnostic problems. Given a set of competing hypotheses as to why an error has occurred, the system would then be able to design critical experiments to determine which hypothesis best explained the student's behavior. Together with SAGE's techniques for discovering the appropriate conditions on operators, this method should lead to a general and robust system for constructing models of students' mathematics strategies. In conclusion, though our research is still in the initial stages, we are confident that it will lead to insights about the nature of student modeling, the nature of strategy learning, and the relation between them.

## Acknowledgements

This research was supported by Contract N00014-83-K-0074 from the Office of Naval Research. I would like to thank Derek Sleeman for discussions that led to many of the ideas presented in this paper.

## References

- Brown, J. S. and Burton, R. R. Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science*, 1978, 2, 155-192.
- Brown, J. S. and VanLehn, K. Repair theory: A generative theory of bugs in procedural skills. *Cognitive Science*, 1980, 4, 379-427.
- Langley, P. Strategy acquisition governed by experimentation. *Proceedings of the European Conference on Artificial Intelligence*, 1982, 171-176.
- Sleeman, D. Inferring (mal) rules from pupils' protocols. *Proceedings of the European Conference on Artificial Intelligence*, 1982, 160-164.
- Young, R. M. and O'Shea, T. Errors in children's subtraction. *Cognitive Science*, 1981, 5, 153-177.

# **Paper Session #7**



# On Association Techniques in Neural Representation Schemes

John A. Barnden  
Computer Science Department  
Indiana University, Bloomington, Indiana.

## Section 1: Introduction

It has often been proposed that, in the brain, associations between information items take the form of suitable settings for synaptic weights [e.g. Anderson and Mozer (1981), Anderson et al (1977), Fahlman (1979, 1981), Feldman (1981), Feldman and Ballard (1982), Goddard (1980), Hebb (1949), Hinton (1981), Kohonen et al (1981), Wickelgren (1979)]. An information item is implemented as a potential or actual pattern of neural activity in some particular set of neurons. An association from item I to item J is implemented as the existence of suitable synaptic weights on neural paths from the neuron set for I to that for J, such that the active presence of the I pattern tends to cause the J pattern to appear. The patterns are anchored, in the sense that the identity of the particular neurons whose activity constitutes a pattern is crucial. In what I shall call the *dedication* approach [Feldman (1981), Feldman and Ballard (1982), Fahlman (1979, 1981), Goddard (1980), Hebb (1949), Wickelgren (1979)], all or many of the neurons in the neuron set for an information item are individually dedicated to that item, in that they do not appear in the neuron sets for other items. (The dedicated neurons are often called "grandmother", "pontifical" or "cardinal" cells.) In what I shall call the *sharing* approach [Anderson and Mozer (1981), Anderson et al (1977), Hinton (1981), Kohonen et al (1981)], individual neurons in the set generally belong to the sets for many other items as well.

We shall look at various problems facing currently proposed schemes which encode association by means of synaptic weight values, when they try to account for *rapid, complex information-processing* such as is involved in understanding and producing natural-language or acting in the world. Some of the schemes do address certain specialized types of short-term processing, but as far as I am aware they do not deal in any general way with the problems to be discussed. We have space here for no more than a brief look at the problems. A more detailed paper on the subject is in preparation.

## Section 2: Some Problems for Synaptic-Weight Schemes

We adopt the working hypothesis that we must show how the neural mass could act as an implementation of semantic network processing of the sort typically postulated in AI and cognitive psychology. We shall assume networks in which relationships as well as non-relational items are coded as nodes, the links being left for restricted "syntactic" uses such as linking a relationship node to the nodes for the partakers in the relationship. We take the nodes to be implemented as neural activity patterns, and we take the links to be implemented as individual associations encoded as synaptic weight settings. These assumptions are in reasonable accord with the approaches taken in Fahlman (1979, 1981), Feldman (1981), Feldman and Ballard (1982), Hebb (1949), Hinton (1981), Kohonen et al (1981), and Wickelgren (1979). The problems on which we focus are association (link) deletion and node marking.

## Network Alteration and Traversal

If semantic network manipulations occur in short-term cognition, presumably they include the traversal of networks and the modification of networks (whether by a change of graph structure or by replacement of one node by another). Such an alteration presumably involves replacement of the piece of net to be altered by a new piece of net which is linked in properly either (a) to the rest of the net as it is or (b) to a copy of the rest of the net. The replaced piece in case (a), or all of the old net in case (b), must somehow be put out of play, whether by being marked, inhibited in some way, or isolated by having associations leading into it deleted. Explicit deletion of an association (link) has received little attention in work on synaptic-weight schemes, and it raises difficulties. Presumably the neural substrate must be put into something like the condition it would have been in had the association not been present. This is no great problem in a dedication scheme, because the synaptic weights defining the association can be reduced to some small value. In a sharing scheme, however, the synaptic weights are important also to other associations, so the particular state they are left in depends crucially on what those associations are. The trouble is that, unless the whole network is somehow traversed or rebuilt, the deletion operation has no guidance as to what those associations are. Similar comments apply to the idea of "inhibiting" a piece of net. By this I mean leaving the structure essentially intact but switching on some special agency which stops the nodes in question from being activated, and/or which stops the associations from being used. The problems in the remaining alternative, marking, receive attention below.

To turn to traversal of networks, it is standard for traversal algorithms implemented on computers to use a marking scheme to ensure that parts of the structure are not traversed more than once. A form of marking would for the same reason be needed in a neural implementation of traversal. To avoid marking we could suggest a random-walk process, which may well duplicate parts of the traversal. This would appear to be a rather error-prone and/or time-consuming technique, and its adoption in a neural model would require convincing argumentation.

## Marking

Fahlman's (1979, 1981) dedication scheme makes heavy use of marking to effect certain types of inference and structure matching. Hinton (1981) uses a marking scheme for structure matching which is more sophisticated but similar in spirit. Marking can be used to get the effect of pointers. If we have some way of activating nodes which are marked in a specific way, then we get the effect of following a pointer.

How is a node marked? We note first that we should be able to provide marking of several different types, and that nodes must be able to be unmarked as well as marked. The first suggestion is that the nature of the potential firing of the individual neurons in the node's neuron-set is changed. For instance, the firing trains of a neuron could have several distinct possible patterns, corresponding to different types of marking. The objection to this is that it involves a major change in the philosophy of synaptic-weight schemes, where it is usually assumed that neurons can be more or less active (e.g. can fire at higher or lower frequency) to indicate the "confidence" with which the node is present, but cannot be active in symbolically distinct ways. Once the door is opened to distinguishing between modes of firing for symbolic purposes, the question arises of whether significant symbolic information of more general sorts should not be encodable in firing patterns. Also, in a sharing scheme, a neuron contributes to several nodes, some of which may be marked and some not.

A second suggestion is that the node to be "marked" is replaced temporarily by a new node which acts as a marked version of the original node. But here we are appealing to an operation of structure alteration as discussed above. As we saw, the replaced node must be put out of play. In a sharing scheme, it seems that the most viable alternative is to put it out of play by marking it! We therefore have a vicious circle. In a dedication scheme, the replaced



node could be inhibited or isolated, but this is a cumbersome process if done merely for the purpose of marking, especially when we consider the possible need for later unmarking of the node.

A third possibility is to have a special node which acts as an explicit mark and is put into association with the node to be marked. Even in the dedication scheme this is an over-elaborate proposal. At one extreme, we have the possibility that there is just one special node, so that all the neural sets implementing nodes which might ever need to be marked have to be connected by a neural path to this special mark node. At the other extreme, we could have a distinct special mark node for every distinct non-mark node. At either extreme, or anywhere in between, a large amount of 'hardware' is set aside just for marking purposes. A sharing scheme faces analogous difficulties, but also faces a particular difficulty in unmarking: the mark, or at least its association to the marked node, must be put out of play, but in the present case we do not want to mark a mark! We might get round this problem by stipulating that the neurons used in the mark node cannot be used for anything except marking, i.e. that they are dedicated to marking; perhaps then the association between mark node and marked node can be easily broken (in view of our comment above that deletion of an association does not appear to be a problem in a dedication scheme). The methodological objection to this is that we are diluting the purity of the philosophy of sharing schemes by letting in dedication in restricted cases for ad hoc reasons.

The last suggestion we make is that extra neurons are somehow included temporarily in the neuron set for the node. We could suppose that each dedicated neuron set has several subsets of special neurons, one subset per marking type. A node is considered to be marked if and only if the appropriate subset of neurons is activated when the main neurons for the node are activated. However, this marking technique is rather ad hoc, since it requires the idea of marking to be built into the very hardware of the system, and we begin to wonder why we should not allow specialized neuron sets to hold symbolic structures more complicated than marks. The sharing schemes face a further difficulty. It is not at all clear how the process of unmarking would work unless the special mark neurons for a node are distinct from the neurons used by any other node; but then we have a restricted form of dedication much as in the third proposal.

A general observation about all the above marking proposals is that they treat marks as data items which need special mechanisms for their implementation or use. This contrasts with marking in computers, where an algorithm may use data items in such a way that we call them marks, but where those data items require no special mechanisms for their implementation or use.

### Section 3: An Alternative

It is of interest that an alternative which avoids the problems of Section 2 can be provided. We shall continue to assume that the issue is the implementation of semantic networks and their manipulation in the course of short-term information-processing. We suppose still that nodes are implemented as patterns of neural activity. However, our patterns are substantially *unanchored* with respect to neural location: their precise positions in the neural mass are irrelevant. To make sense of this, we suppose that the neural areas in which the patterns reside are regular in structure. In fact, we make the tentative, simplifying assumption that the areas are physically structured as arrays (typically of dimension two). We call each such area a *pattern matrix* (PM). A PM is an array of neural circuits called PM elements. Each PM element can be active in any one of a certain number (say a dozen or two) modes, any combination of modes being allowed. Short-term information structures are patterns of PM activity over the PMs. The pattern in a PM is generally composed of well-defined subpatterns, some of which play the rôle of network nodes. The precise position of a subpattern is unimportant, although its position relative to other subpatterns may be crucial, as we shall see in a moment.

A distinctive feature of our scheme is the way in which subpatterns can be in association so as to form structures. The association does not take the form of transmission-facilitated neural paths. Instead, association is by *adjacency* and *similarity*. Adjacency association is similar to the association of data items in a computer by virtue of their being in adjacent locations. So, two subpatterns in a PM which are adjacent to each other may be taken to be associated. (As a special case, one of the subpatterns may be like a closed boundary and contain the other subpattern.) Similarity association is akin to content addressing in computers. There is a mechanism attached to PMs such that the presence of a subpattern in a PM can cause sufficiently similar subpatterns in this and other PMs to be be "highlighted" by high activity in some mode (whose identity is passed to the mechanism as a parameter). Subpatterns which are thus associated by similarity can be placed adjacent to other subpatterns (e.g. nodes) which they can be considered to "label". The labelled subpatterns can thereby be regarded as being indirectly associated by a combination of adjacency and similarity association. A further form of association is derived from adjacency association: two subpatterns in different parts of a PM can be joined by a line-like supattern (whose ends are adjacent to the first two subpatterns). Such line-like subpatterns are analogous to the link lines in a diagram of a semantic network. They are also analogous to the neural paths in synaptic-weight schemes, but instead of a facilitated transmission path there is a path of activated neural networks (PM elements).

We suppose that connected to the PMs there is a neural mechanism embodying a production system. The condition part of a rule responds to the presence of fairly simple combinations of primitive subpatterns. The primitive subpatterns in the case of network-like patterns would be subpatterns acting as nodes, links and labels (or perhaps components of labels). The action part of a rule is able to insert subpatterns, delete subpatterns, copy subpatterns, move subpatterns around in PMs, follow link-like subpatterns, highlight subpatterns, invoke the pattern-similarity association mechanism and communicate with mechanisms outside the PM production system. The set of rules does not change in the short term. We envisage the rules to be implemented as neural networks attached to the PMs, and the triggering of rules to occur by just the sort of neural associative techniques as are proposed in synaptic weight schemes. Also, patterns in PMs could associate to long-term structures by such mechanisms. It should be noted that subpatterns in PMs are not themselves the full embodiments of concepts or other entities; rather they are merely symbols for or "ambassadors" of those entities.

In our scheme all information in a semantic network, including association, is encoded as activity patterns. This provides greater uniformity and elegance than is present in synaptic-weight schemes, in which there are two completely different embodiments of information: (potential) patterns of activity and synaptic weights. At the same time, by being relatively close to the way computers work (if we take adjacency, similarity and line-linking to correspond to locational adjacency, content addressing and direct addressing in computers) our proposal has the advantage that techniques developed for information-processing in computers can relatively easily and plausibly be supposed to occur in the brain. For instance, nodes and links can be labelled and marked very simply and naturally, either by highlighting them or by placing label subpatterns next to them. No special extra mechanisms need be postulated, and our deeming a particular feature of a PM pattern to be a mark is merely a result of the way particular rules use the feature (cf. the comments on marking in computers at the end of Section 2). Subpatterns which are associated by adjacency, line-linking or label-similarity can be altered without affecting their inter-association, because the patterns embodying the associations are independent of the patterns associated. Associations can easily be deleted, whether by moving subpatterns so that they are no longer adjacent or by removing labels or line links. (Such removal is performed by suppressing the activity in the PM elements concerned.) Therefore, the problems of deletion and marking that we noted in Section 2 do not arise for our scheme, so that operations such as traversal no longer present special difficulties.

There is no space here to present the scheme in more detail. One version of the scheme is reported in Barnden (1982a, 1982b). Barnden (1982b) goes into considerable detail concerning the possible operation of the production system in manipulating network-like patterns in PMs. The scheme appears to be no less well supported by known facts about the brain than are

the synaptic-weight schemes – indeed, by virtue of its foundation on arrays of neural nets it fits in more naturally with the regular and ubiquitous columnar organization of cortex [Mountcastle (1978)] than those schemes do.

## References

- Anderson, J.A. and Mozer, M.C. Categorization and selective neurons. In Hinton and Anderson (1981).
- Anderson, J.A., Silverstein, J.W., Ritz, S.A. and Jones R.S. Distinctive features, categorical perception, and probability learning: some applications of a neural model. *Psychological Review*, 1977, 84, 413-451.
- Barnden, J.A. The integrated implementation of imaginal and propositional data structures in the brain. *Procs. 4th. Annual Conf. of the Cognitive Science Society*, Ann Arbor, Michigan, 1982a.
- Barnden, J.A. A continuum of diagrammatic data structures in human cognition. Tech. Rep. 131, Computer Science Dept., Indiana University, October 1982b.
- Fahlman, S.E. *NETL: A system for representing and using real-world knowledge*. Cambridge, Mass.: MIT Press, 1979.
- Fahlman, S.E. Representing implicit knowledge. In Hinton and Anderson (1981).
- Feldman, J.A. A connectionist model of visual memory. In Hinton and Anderson (1981).
- Feldman, J.A. and Ballard, D.H. Connectionist models and their properties. *Cognitive Science*, 1982, 6, 205-254.
- Goddard, G.V. Component properties of the memory machine: Hebb revisited. In P.W. Jusczyk and R.M. Klein (Eds.), *The nature of thought: Essays in honor of D.O. Hebb*. Hillsdale, N.J.: Lawrence Erlbaum, 1980.
- Hebb, D.O. *Organization of behaviour*. New York: Wiley, 1949.
- Hinton, G.E. Implementing semantic networks in parallel hardware. In Hinton and Anderson (1981).
- Hinton, G.E. and Anderson, J.A. (eds) *Parallel models of associative memory*. Hillsdale, NJ: Lawrence Erlbaum, 1981.
- Kohonen, T., Oja, E. and Lehtiö, P. Storage and processing of information in distributed associative memory systems. In Hinton and Anderson (1981).
- Mountcastle, V.B. An organizing principle for cerebral function: the unit module and the distributed system. In G.M. Edelman and V.B. Mountcastle, *The mindful brain*. Cambridge, Mass.: MIT Press, 1978.
- Wickelgren, W.A. Chunking and consolidation. *Psychological Review*, 1979, 86, 44-60.



George Graham & G. Lynn Stephens  
Department of Philosophy  
University of Alabama in Birmingham  
Birmingham, Alabama 35294

## PAINS AND STRONG COGNITIVISM

Strong Cognitivism in psychology and philosophy is roughly the position that all and only cognitive states and processes (propositional attitudes) are psychological. In philosophy, Strong Cognitivism is an essential feature of the regnant philosophy of mind, Computer Functionalism (Dennett, 1978). The exact characterization of a cognitive state or process is a matter of controversy (Dretske, 1981). But the issue does not have to be settled here. Suffice it to say, beliefs and desires are paradigm cognitive states; pains -- where 'pains' is understood to refer to immediate felt qualities which are independent of any propositional content or representational role -- are paradigm noncognitive states.

It is sometimes said (e.g., Block, 1978) that pains are a threat to Computer Functionalism; that they are noncognitive and hence incompatible with Strong Cognitivism.

But the conclusion that pains are incompatible with Strong Cognitivism, when drawn from the premiss that pains are noncognitive is a nonsequitur. Pains are incompatible with Strong Cognitivism only if they are also psychological. If pains are not psychological -- if having a pain is not being in a psychological state -- then that pains are noncognitive does not threaten Strong Cognitivism or Computer Functionalism. To bring this out, compare noticing a pain with seeing a tree. Strong Cognitivism is not required to account for trees. It may need to account for tree-perception. But it doesn't need to explain what a tree is. That is the business of some nonpsychology, because trees are nonpsychological sorts of things.

Now if the same point is true of pain, which is to say, if pains are not psychological, then Strong Cognitivism -- and Computer Functionalism -- does not

have to account for pain. It may need to explain some of the causes and effects of pain, such as, e.g., the desire to be free of pain. But it does not have to characterize what pain is. That is not the business of psychology.

We wish to argue that pain is not psychological. In fact, we believe that pain is self-evidently nonpsychological. To bring this out, consider the phenomenology of pain. When people are in pain it is always some part of the body that "hurts". People speak of feeling or noticing pains in necks, toes, heads, and so forth. "There is a burning sensation in my lower back." "My throat is sore!" "I have a prickly feeling behind my left knee." Further, pain is often spoken of as moving or spreading from one bodily location to another. "The pain starts in my hip and radiates down the side of my leg." A natural or manifest interpretation of such locutions is that pain occurs in the body, not in the "mind." Those who deny that the throbbing pain I feel in my big toe is actually in my toe deny that things are as they evidently appear. They say, "You do not have a pain in your toe. You have a toe-pain. What you call 'a pain in your toe' is actually a state of mind." However it should be noted that this sort of response is unnatural, and arrived at only by reflection. Unaided, or unprejudiced, by ideologies people are inclined to say that pains are in their bodies and not in their "minds". Why say otherwise?

One hurdle to the thesis that pains are not psychological is the phenomenon of phantom limb pain; the idea being that phantom limb pains show that people can be mistaken about where pains are and that thus the phenomenology of pain should not be trusted. Pains are in minds, not bodies.

To infer from phantom limb pains that pains are in minds rather than bodies is to accuse victims of phantom limb pain not merely of error but gross error. It is to say that they are mistaken not only in thinking of pain as in their limbs, but in thinking of pain as in their bodies. It would be more reasonable, we believe, to accuse victims of phantom limb pain of mere error; of thinking of pain as in a certain



place in their body (the absent limb) when in fact it<sup>+</sup> is in another place in their body, and not in their "mind". As such, the lesson of phantom limb pains is not that pains are psychological. It is that people can bodily mislocate their pains.

Consider the following analogy. Boaters sometimes report that oars are bent when submerged in water. To infer from such illusions that the oars are not only not bent but not in water is to accuse the boaters of gross error. In contrast, to infer only that the oars are not bent is to accuse the boaters of mere error. Which hypothesis is more reasonable? The hypothesis of mere error, of course: the oars though in water are not bent. By analogy: the pain though in the body is not in the phantom limb.


Another hurdle is the privacy of pain: the idea being that pain is private and that whatever is thus private is psychological.

We contend that pain is not private; that some of the psychological states associated with pain might be private, but that pain itself is not a private sort of thing. As an immediate felt quality, pain is a universal, capable of multiple instantiations or instances. I can know how your pain feels, because I may have had the same feeling, the same pain, yesterday. And again, even if certain psychological states associated with pain are private, this would not make pain private. Suppose I am thinking of Moscow. My thought might be private; but Moscow is not. Analogously, suppose I perceive a sharp, stabbing pain in my left knee. My perception might be private; but the pain is not.

We don't deny that certain psychological states associated with pain are or can be private. We don't deny that the issue of the privacy of pain is complicated by the possible privacy of the states associated with pain. We simply deny that pain is private; and thus that pain is psychological because it is private.

The final hurdle to the thesis that pain is not psychological is the notion that pain is cognitive state-dependent; that someone cannot be in pain without, e.g.,

believing it. This fact -- the cognitive state-dependency of pain -- is often thought to make pain psychological. But it doesn't. Coins, e.g., are cognitive state-dependent but not psychological. Something isn't a coin unless, e.g., people believe that it is legal tender. And it is of course possible to deepen the analogy between pains and coins if we assume that pains are bodily: standing in relation to a cognitive state (e.g., the belief that it is legal tender) may be necessary for a certain piece of metal to be a coin, just as standing in relation to a cognitive state (e.g., the desire to be free of it) may be necessary for a certain bodily state to be a pain. About the only thing that cognitive state dependence proves is that pains are had only by creatures with cognitive states, and this isn't enough to show that pains are psychological.

One last point. Some anti-functionalists believe that Strong Cognitivism can be trumped by imagining a creature fully endowed with cognitive states (including such states as the belief that his toe hurts, etc.) but absent pain. What this is supposed to prove -- philosophers will recognize the point as the Absent Qualia Objection to functionalism -- is that pain is not a cognitive state and that functionalism is therefore incomplete as an account of the psychological. But what this means to us is that advocates of the Objection are guilty of a logical error. If it's possible for a fully endowed cognizer to be absent pain, why should this show that functionalism is defeated. It needs to be shown in addition that pain is psychological. And it's not. The metaphysical commitments of Strong Cognitivism -- as well as of Computer Functionalism -- are not threatened by pain. 

Nor is the issue simply one of metaphysics. Tons of research monies have been spent in the search for a psychological conception of pain. Cognitive science is party to this practice. It is of course possible that cognitive science will tell us a lot about certain causes and effects of pain. On the other hand, if we are right that pain is not psychological, the project of a psychological characterization

of pain is doomed to failure.

#### NOTE

The single authorial voice is sometimes used as a stylistic device in this paper, and does not reflect anything substantive about its composition; neither does the order of authorship, which is simply alphabetical.

#### REFERENCES

- Block, Ned. Troubles with functionalism. In C. W. Savage (Ed.),  
Perception and Cognition. Minneapolis: University of Minnesota, 1978
- Dennett, D. C. Brainstorms. Bradford, Vt.: MIT Press, 1978.
- Dretske, Fred. Knowledge and the flow of information. Bradford, Vt. :  
MIT Press, 1981.



# ANALYZING COOPERATIVE COMPUTATION

Geoffrey E. Hinton  
Computer Science Department  
Carnegie-Mellon University  
Pittsburgh, Pennsylvania 15213

Terrence J. Sejnowski  
Biophysics Department  
The Johns Hopkins University

## ABSTRACT

Making a perceptual interpretation can be viewed as a computational process in which a plausible combination is chosen from among a large set of interdependent hypotheses. In a cooperative computation the hypotheses are implemented by units that interact non-linearly and in parallel via excitatory and inhibitory links (Julesz, 1971; Marr & Poggio, 1976; Sejnowski, 1976). A particular perceptual task is specified by external inputs to some of the units and the whole system must then discover a stable state of activity in which the active units represent the hypotheses that are taken as true. We describe a search procedure based on statistical mechanics that finds near optimal combinations of hypotheses with high probability, and we show that the hardware units required for its efficient implementation are similar to neurons. Even though the individual units are non-linear, there is a linear relationship between the synaptic weights and the logarithms of the probabilities of global states into which the system settles. This makes it possible to implement a convergent learning procedure which specifies just how the synaptic weights need to be changed in order to learn the constraints in a given domain.

## Introduction

Consider the problem of making a 3-D interpretation of a 2-D line drawing. Each line in the picture, considered in isolation, could depict any one of a large set of 3-D edges. People resolve this local ambiguity by using assumptions about the ways in which edges go together in the 3-D world. These assumptions make some combinations of edges far more plausible than others. There are two roughly separable problems in understanding the use of assumptions in perception. The first is to specify clearly what the assumptions are, and the second is to find a search procedure that can discover interpretations which optimally fit the input data and the assumptions, even when some of the assumptions conflict with one another (Attneave 1982). Our concern here is with the second problem: How can we discover interpretations that optimally fit a large set of plausible assumptions?

Attneave (1982) and others (Hinton 1977) have proposed cooperative models in which neuron-like hardware units represent particular 3-D edges and the rules are implemented by excitatory and inhibitory interactions between these units. Each line in the drawing provides input to the whole set of 3-D edges which are consistent with it, and under the influence of this input the whole system settles into a stable state of activity which represents the interpretation. It is not obvious that such a search process can be made to work. The apparent difficulty of analyzing the behaviour of cross-coupled, non-linear systems makes it tempting to believe that the *only* way to make progress is through computer simulation. In this paper we attempt to show that mathematical analysis is possible and illuminating.

Most of the existing proposals for cooperative search mechanisms assume that there are real-valued activity levels which change smoothly during the search (Rosenfeld, Hummel & Zucker, 1976). These activity levels are often associated with the firing rates of neurons, and they are normally used to represent the value of a physical parameter such as slope in depth, or the current probability that a hypothesis is correct. The method we shall describe uses a very different representation. The units that stand for hypotheses only have two states, true and false. However, the decision rule which determines which state they enter is probabilistic, so they can change their state even if they are receiving constant input. The use of a probabilistic decision rule makes the cooperative search *easier* to analyze than with a deterministic rule because it makes it possible to apply methods from statistical mechanics. Instead of being a drawback, the non-determinism has the advantage of allowing the system to escape from sub-optimal states. We start by describing a system in which there is a deterministic decision rule that is applied at random moments and then we generalize this case to a non-deterministic rule.

## Cooperative search with deterministic binary units

Hopfield (1982) postulates a system with a large number of binary units. The units are *reciprocally* connected, with the

strength of the connection being the same in both directions. Given the current inputs from outside the system, any particular state of the system has an associated "energy" and the whole system behaves in such a way as to minimize its energy. The energy of a state can be interpreted as the extent to which it violates a set of plausible constraints, so in minimizing its energy it is maximizing the extent to which it satisfies the constraints.

The total energy of the system is defined as

$$E = -1/2 \sum_{ij} w_{ij} s_i s_j - \sum_i (\eta_i - \theta_i) s_i \quad (1)$$

where  $\eta_i$  is the external input to the  $i^{th}$  unit,  $w_{ij}$  is the strength of connection (synaptic weight) from the  $j^{th}$  to the  $i^{th}$  unit,  $s_i$  is a boolean truth value (0 or 1), and  $\theta_i$  is a threshold.

A simple algorithm for finding a combination of truth values that is a *local* minimum is to switch each hypothesis into whichever of its two states yields the lower total energy given the current states of the other hypotheses. If hardware units make their decisions asynchronously, and if transmission times are negligible, then the system always settles into a local energy minimum. Because the connections are symmetrical, the difference between the energy of the whole system with the  $k^{th}$  hypothesis false and its energy with the  $k^{th}$  hypothesis true can be determined locally by the  $k^{th}$  unit (Hopfield, 1982), and is just

$$\Delta E_K = \sum_i (w_{ki} s_i) + \eta_k - \theta_k \quad (2)$$

Therefore, the rule for minimizing the energy contributed by a unit is to adopt the true state if its total input exceeds its threshold, which is the familiar rule for binary threshold units (Minsky & Papert, 1968).

## Using probabilistic decisions to escape from local minima

The deterministic algorithm suffers from the standard weakness of gradient descent methods: It gets stuck at *local* minima that are not globally optimal. This is an inevitable consequence of only allowing jumps to states of lower energy. If, however, jumps to higher energy states occasionally occur, it is possible to break out of local minima. An algorithm with this property was introduced by Metropolis *et. al.* (1953) to study average properties of

thermodynamic systems (Binder, 1978) and has recently been applied to problems of constraint satisfaction (Kirkpatrick, Gelatt & Vecchi, in press). We adopt a form of the Metropolis algorithm that is suitable for parallel computation: If the energy gap between the true and false states of the  $k^{th}$  unit is  $\Delta E_k$  then regardless of the previous state set  $s_k = 1$  with probability

$$p_k = \frac{1}{(1 + e^{-\Delta E_k/T})} \quad (3)$$

where  $T$  is a parameter which acts like temperature (see fig. 1). This parallel algorithm ensures that in thermal equilibrium the relative probability of two global states is determined solely by their energy difference, and follows a Boltzmann distribution.

$$\frac{P_\alpha}{P_\beta} = e^{-(E_\alpha - E_\beta)/T} \quad (4)$$

At low temperatures there is a strong bias in favor of states with low energy, but the time required to reach equilibrium may be long. At higher temperatures the bias is not so favorable but equilibrium is reached faster.

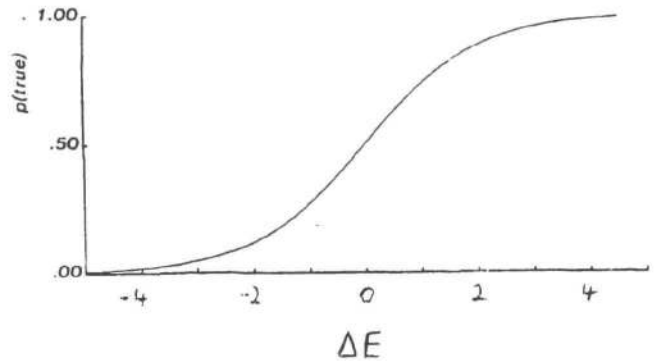


Figure 1

Probability  $p(\Delta E)$  that a unit is in its "true" state as a function of its energy gap  $\Delta E$  plotted for  $T=1$  (Eq. 3). As the temperature is lowered to zero the sigmoid approaches a step function.



## Reducing the time to reach equilibrium

One technique that can be used to reach a good equilibrium distribution quickly is to start at a high temperature and then to cool down (Kirkpatrick *et. al.*, in press). This type of search by "simulated annealing" initially finds a large-scale minimum but fluctuates around it because of the high temperature. As the temperature is reduced, a good minimum will be found within the large-scale minimum, and so on. In general, it is impossible to *guarantee* that a global minimum will be found, but a nearly global minimum can be found with high probability.

We are investigating an additional technique which we shall only mention here. Energy barriers are what prevent a system from reaching equilibrium rapidly at low temperature, and if they can be temporarily suppressed, equilibrium can be achieved rapidly at a temperature at which the distribution strongly favors the lower minima. The energy barriers cannot be permanently removed, because they correspond to states that violate the constraints, and the energies of these states must be kept high to prevent the system from settling into them. However, for special cases it is possible to design units which are active during the search process but are quiescent in the final state. When one of these special units is active it lowers the energy of a state that would have been an energy barrier between two local minima. The special units are a way of implementing heuristic knowledge about how to search the space. They have no effect on the energies of final states, and in this respect they are like catalysts.

## Learning

So far, we have assumed that the interactions between the units implement the correct constraints, and we have focussed on the search problem. However, in a system where the weights represent many plausible assumptions that interact, it is not obvious how to choose the weights to produce the desired behavior. We will show that, as a consequence of the probabilistic decision rule, it is possible for a cooperative module to internalize the constraints in any domain simply by being told whether the solutions it settles into are right or wrong. When the module settles to the wrong solution, it modifies the weights so as to raise the energy of that state and thus make it less likely to be found in future. Similarly, good solutions that are not found often enough have their energies lowered when they are found. This simple procedure is effective because of the *linear* relationship between the synaptic weights and the logs of

probabilities of whole states at thermal equilibrium. If we temporarily ignore the thresholds and the external inputs to the units and assume a temperature of 1, we have:

$$\begin{aligned} \ln\left(\frac{P_\alpha}{P_\beta}\right) &= -(E_\alpha - E_\beta) \\ &= \sum_{ij} w_{ij} h_{ij}^{\alpha\beta} \end{aligned} \quad (5)$$

where

$$h_{ij}^{\alpha\beta} = s_i^\alpha s_j^\alpha - s_i^\beta s_j^\beta$$

and  $s_i^\alpha$  is the state of the  $i^{th}$  unit in the  $\alpha^{th}$  global state.

To explain the learning procedure, we invent a hypothetical ideal system which settles into global states with exactly the probabilities required. We then show that if the actual system is told whether its current probabilities for particular states are too high or too low, it can modify its weights so that they more closely resemble the weights in the hypothetical ideal system.

Suppose that under the influence of a constant external input vector, the actual system settles into two different states,  $S_\alpha, S_\beta$  with probability ratio  $P_\alpha/P_\beta$ . Suppose that the probability ratio demanded by the evaluation function (and achieved by the ideal system) is  $P'_\alpha/P'_\beta$  which is higher. The actual system can increase its probability ratio by increasing the energy difference,  $E_\beta - E_\alpha$ . This can be done by adding  $\delta$  to each weight between a pair of active units in  $S_\alpha$  and subtracting  $\delta$  from each weight between a pair of active units in  $S_\beta$ . The net change in a weight is then  $\delta \cdot h_{ij}^{\alpha\beta}$ . We now prove that, provided  $\delta$  is sufficiently small, each application of this learning procedure is guaranteed to reduce the Euclidean distance,  $D$ , between the current set of weights,  $w_{ij}$ , and the ideal ones,  $w'_{ij}$ . Assume that the actual and ideal systems have the same external inputs and thresholds, and that  $T = 1$ . If the error,  $r$ , in the probability ratio achieved by the actual system is

$$r = \ln\left(\frac{P'_\alpha}{P'_\beta}\right) - \ln\left(\frac{P_\alpha}{P_\beta}\right)$$

then from equations 4 and 5, we have:

$$r = -(E'_\alpha - E'_\beta) + (E_\alpha - E_\beta)$$

$$= \sum_{ij} h_{ij}^{\alpha\beta} w'_{ij} - \sum_{ij} h_{ij}^{\alpha\beta} w_{ij}$$

Before applying the learning rule we have

$$D_{before}^2 = \sum_{ij} (w_{ij} - w'_{ij})^2$$

and afterwards

$$\begin{aligned} D_{after}^2 &= \sum_{ij} (w_{ij} + \delta h_{ij}^{\alpha\beta} - w'_{ij})^2 \\ &= D_{before}^2 - \delta \sum_{ij} (2h_{ij}^{\alpha\beta} w'_{ij} - 2h_{ij}^{\alpha\beta} w_{ij} - \delta (h_{ij}^{\alpha\beta})^2) \\ &= D_{before}^2 - \delta(2r - \delta \sum_{ij} (h_{ij}^{\alpha\beta})^2) \\ &= D_{before}^2 - \delta(2r - \delta n) \end{aligned}$$

So the distance is reduced iff  $\delta < 2r/n$

where  $n = \sum_{ij} (h_{ij}^{\alpha\beta})^2$  = the number of weights that are changed.

Having a simple convergent learning procedure for a non-linear system is important because it allows the synaptic weights that implement the energy function to be determined by feedback from the correctness of the interpretation that the system settles into. Thus the constraints implicit in the task can be programmed into the system simply by telling it how well it is doing.

The learning procedure assumes that the system receives feedback from an evaluator that tells it whether the current value of  $\ln(P/P_\beta)$  is greater or less than the ideal value  $\ln(P'_\alpha/P'_\beta)$ . This places a very stringent requirement on the evaluator since it must know about the desired probabilities of whole global states like  $S_\alpha$ . To build these desired probabilities into the evaluator, the representations that the system should use must be decided in advance. A less omniscient evaluator would only know what *some* of the units should do for each input vector and would leave the system to decide for itself how to use the remaining, "hidden" units to achieve this. Suppose, for example, that there is a set of global states  $\Omega_\alpha$  which only differ from one

another in the hidden units that the evaluator cannot see. The evaluator specifies required probabilities of the form:

$$P'_{\Omega_\alpha} = \sum_{\alpha \in \Omega_\alpha} P_\alpha$$

but it does not specify how the total probability should be distributed over the various states in  $\Omega_\alpha$ . The different ways of distributing the probability correspond to using different representations in the hidden units.

If there are units that are hidden from the evaluator, it is impossible to define a single hypothetical ideal set of weights. There may be many different complete sets of weights which would yield the required behaviour for the "visible" units, and these sets do not, in general, form a convex set. In travelling towards one suitable set of weights, the system may travel away from other equally suitable sets, so convergence on any one set is not guaranteed. This means we need a different measure of the progress of learning in order to prove convergence. A suitable measure is the information theoretic distance,  $G$ , between the actual and required probability distributions over all  $2^n$  states of the  $n$  visible units:

$$G = \sum_{\alpha} P_{\Omega_\alpha} \ln \left( \frac{P_{\Omega_\alpha}}{P'_{\Omega_\alpha}} \right)$$

The value for  $G$  depends implicitly on the  $w_{ij}$  and so  $G$  can be reduced by changing each weight by an amount that is proportional to the partial derivative of  $G$  with respect to that weight. We describe this learning rule further in Hinton and Sejnowski (1983). It is guaranteed to find a minimum of  $G$ , but it may only be a local minimum rather than a global one. Local minima occur when the system is doing the best that it can given the representations it has learnt in the hidden units. To do better it has to change these representations which involves a temporary setback in how well it meets the requirements on the probabilities of the states of the visible units. Of course, if the modifications to the weights are probabilistic so that  $G$  can sometimes increase, it is possible to escape from local minima and ensure that after enough learning there is a bias in favor of the better local minima.

## Relation to the brain

There are two different ways to interpret the input-output function that hardware units should have to implement the parallel search (Fig. 1). During a short interval the sigmoid curve describes the probability of a unit being in the true state as a function of the energy gap between the false and true states. For much longer time intervals the curve

describes the proportion of time that the unit is in its true state. If we assume that a hypothesis which is true all the time is represented by a neuron firing at its maximum rate, then the curve in Fig. 1 can be interpreted as the firing rate of a neuron as a function of its average input (Sejnowski, 1977). However, the way in which truth values are represented by action potentials is not the kind of simple encoding in which two different voltage levels stand for the two truth values. Instead, it appears that an action potential only provides a delta-function type of signal that drives integrative processes in the recipient neurons. This amounts to treating a hypothesis as "true" for a whole refractory period after an action potential has been emitted.

The parallel algorithm for cooperative search depends on the computation of energy gaps  $\Delta E_i$ . In the case of symmetrically connected units the global energy gaps can be computed locally by single units. It seems unlikely that neurons in cerebral cortex are symmetrically connected, but if a neuron receives many inputs it can still estimate what its contribution to the total energy would be if all the connections had been symmetrical. In simulations, asymmetric networks behave like symmetric ones with added noise (Hopfield, 1982), and time delays in transmission have a similar effect. Provided that the task requires symmetric connections, as is the case for problems of constraint satisfaction, an asymmetric network can closely approximate the performance of a symmetric one.

The computational model analyzed in this paper is not a realistic model of processing in cerebral cortex, for it falls far short of explaining the known anatomical and physiological facts. The analysis may, however, provide insight into a class of computational devices that depend on probabilistic parallel processing. Understanding general properties of this class may be a useful first step in understanding particular highly-evolved members of the class. For example, the probabilistic nature of electrical responses of single neurons is well-known, but has generally been regarded as evidence of imprecision. Probability, however, may be a central design principle of cerebral cortex (Sejnowski, 1981). A very close approximation to the function in Fig. 1 can be implemented by simply adding Gaussian noise to a binary threshold unit, with the standard deviation of the noise acting like temperature. We suggest that fluctuations may be deliberately added to neural signals to avoid locking the network into unwanted local optima and to provide the linear conditions needed for efficient learning. The issue of noise in the nervous system deserves renewed experimental investigation and further theoretical analysis.

## Acknowledgements

This work was supported by grants from the System Development Foundation and by earlier grants from the Sloan Foundation to Don Norman and to Jerry Feldman. We thank Francis Crick, Scott Fahlman, David Rumelhart, and Paul Smolensky, for helpful discussions.

## REFERENCES

- Attneave, F. Pragnanz and soap-bubble systems: A theoretical exploration. In J. Beck (Ed.) *Organization and Representation in Perception*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1982.
- Binder, K. (Ed.) *The Monte-Carlo Method in Statistical Physics* New York: Springer-Verlag, 1978.
- Hinton, G. E. Relaxation and its role in vision. PhD Thesis, University of Edinburgh, 1977; Described in: *Computer Vision*, D. H. Ballard & C. M. Brown (Eds.) Englewood Cliffs, NJ: Prentice-Hall, 1982, pp. 408-430.
- Hinton, G. E. & Sejnowski, T. J. Optimal perceptual inference. To appear in: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, Washington DC, June 1983.
- Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences USA*, 1982, 79 pp 2554-2558.
- Julesz, B. *Foundations of Cyclopean Perception* Chicago: University of Chicago Press, 1971.
- Kirkpatrick, S. Gelatt, C. D. & Vecchi, M. P. Optimization by simulated annealing. *Science* (in press)
- Marr, D. & Poggio, T. Cooperative computation of stereo disparity. *Science*, 1976 194, p 283-287.
- Metropolis, N. Rosenbluth, A. W. Rosenbluth, M. N. Teller, A. H. Teller, E. *Journal of Chemical Physics*, 1953 6, p 1087.
- Minsky, M. Papert, S. *Perceptrons* Cambridge, MA: MIT Press, 1968.
- Rosenfeld, A. Hummel, R. A. & Zucker, S. W. Scene labeling by relaxation operations. *IEEE Transactions on Systems, Man, & Cybernetics*. SMC-6, 1976, pp 420-433.
- Sejnowski, T. J. On global properties of neuronal interaction. *Biological Cybernetics*, 1976, 22, pp 85-95.
- Sejnowski, T. J. Storing covariance with non-linearly interacting neurons. *Journal of Mathematical Biology*, 1977, 4, pp 303-321.
- Sejnowski, T. J. Skeleton filters in the brain. In G. E. Hinton & J. A. Anderson (Eds.) *Parallel Models of Associative Memory*. Hillsdale, NJ: Erlbaum, 1981, pp 189-212.



# **Paper Session #8**





# Long Term Recency Effects In Recalling Previous Answers

M.W. Lansdale  
Systems Psychology Department  
ITT Europe \*

## ABSTRACT

Successive tests of the same memory often appear to change it, either by consolidation of a correct response, or an increased probability of repeating errors. This raises the question as to whether the act of attempted recall modifies existing memory traces or creates new ones. Experimental work in this area is complicated by difficulties (a) in isolating what the interactions between successive trials actually reflect, and (b) determining exactly when such interactions have taken place.

A methodology is described in which these difficulties can be overcome within the domain of a specific experimental paradigm: the repeated cueing (RTTT) method used by Jones (1974). Analysis provides three significant findings. Firstly, attempts at recall produce an engram which is normally unavailable after a small number of subsequent trials. Secondly, in a small number of cases these traces may survive for at least the duration of the experiment. Finally, analysis of subject's confidence ratings indicates that subjects are able to distinguish between recalls of stimuli and previous answers, implying an addition to, rather than a corruption of, existing memory.

## INTRODUCTION

Many of the techniques we use for studying memory require that the recall be tested more than once. However, it is becoming increasingly likely that repeated tests do not leave memory unchanged for later trials: evidence is accumulating to suggest that attempts at recognition or recall can themselves give rise to traces which may compete for recall with the original memory (e.g. Kay, 1955), or in some way alter the probability of recall (Izawa, 1970).

This has a number of important consequences. Firstly it represents a source of interference in experimental data which could lead to misleading inferences. Tests of all-or-none recall, for example, look at the consistencies in recall of a stimulus from one trial to the next. However, if the subject is merely recalling his previous responses, then all-or-none patterns in recall will predominate and may not represent the state of memory for the original stimulus. Inter-response interactions are therefore of theoretical interest in their own right.

Secondly, from an applied point of view, the interrogation of memory is of everyday importance in society: law being the most obvious case. Here it is important that the accuracy of recall is preserved as far as possible. The work of Loftus and others (e.g. Loftus, Miller & Burns, 1978) has supported the belief that the use of misinformation or leading questions can corrupt a witness' recall. Legal protocol has evolved to avoid this as far as possible, but since the repeated questioning of a witness' memory is unavoidable (not least by the witness himself), it is still important to know and understand how memory may be changed by attempts at recall.

One difficulty in the experimental study of this process is that in most experiments the recalls of previous responses are indistinguishable from guesses or recalls of the original stimulus. This is because a single response only is required, as in paired associate experiments (e.g. see Izawa 1970). Such experiments give little information as to the interactions between responses.

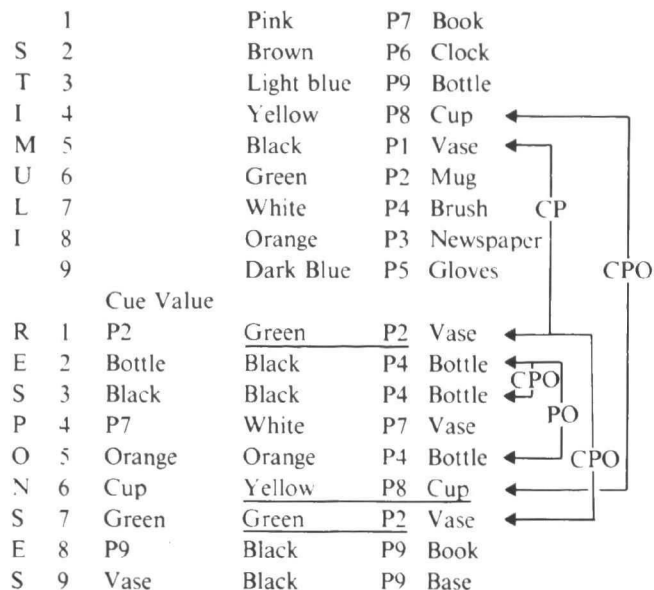
The aim of this research is to study the recall of previous answers in an experiment in which this distinction can be made operationally. This experiment turns on the repeated cueing of multicomponent stimuli after the manner devised by Jones (1974) to test his Fragmentation Hypothesis. Although this is not central to the study of sequential effects in memory, it nevertheless provides the basis for studying them in this case. Since this is a relatively new idea, it is therefore useful to summarise Jones' work briefly before describing the current research in more detail.

The Fragmentation Hypothesis states that memory of a stimulus (such as a picture) is equivalent to a fragment of the original stimulus (in practice a subset of its attributes), structured in such a way that recall of the entire fragment occurs in an all-or-none manner when and only when the cue for recall is contained in the fragment. The critical test of this hypothesis depends on the repeated cueing of a multicomponent memory by each of the stimulus attributes in turn without providing any feedback of results. If the hypothesis holds, the patterns of recall should unambiguously correspond to one or another of the possible fragmentations of the stimulus.

Two experiments by Lansdale (1979) showed that this unusually precise hypothesis holds very well. The first of these is of interest here. The stimuli used were scenes of a billiard table in which three dimensions were defined: a white object (O) on the side of the table, a random pattern of red balls (P), and a coloured ball (C). Nine alternative values of each attribute were used and the stimulus set consisted of an orthogonal set of nine such pictures. With each attribute value being used once as cue, the data from each presentation set consists of a sequence of nine stimulus combinations of C, P and O values followed by 27 test trial combinations, in which one of C, P, or O is given as cue and

the other two are the subject's responses. A hypothetical sequence of stimuli and responses is given in Figure 1, with those elements of the response corresponding to identifiable fragments underlined.

**Figure 1: Relationships between stimuli and responses.**



It can be seen how in this type of data recall of previous responses can be recognised. They will appear as congruences between response combinations (e.g. between trials 2 and 3 or trials 1 and 7) over and above that normally expected by correct recall or chance repetition of errors. The possible effect of recalling previous errors is apparent if one considers the possibility that the subject guesses 'green' correctly to the cue "P2" at trial 1 and then repeats this response at trial 7. The patterns of 'recall' indicate that cue "P2" elicits recall of "green" and vice-versa, from which it would be inferred incorrectly that a [CP] memory fragment exists. Clearly repetition can therefore misrepresent the data in suggesting an all-or-none memory where no memory need exist or need not be all-or-none in reality.

From this brief discussion of the experimental technique, it can be seen that the means exist to investigate the possible effects of answer repetition more deeply; and that this also has great significance for the validity of the repeated-cueing of memory. To do this, it is first necessary to show that it occurs significantly more often than would normally be expected by chance, and to this end a sequential analysis, described in the next section was carried out.

## SEQUENTIAL ANALYSIS

In investigating repetitions of previous answers in these data there are two questions to be resolved:

- How does one define a response as a recall of a previous answer?

- How does one show that this is more than just a chance event?

## Defining Previous Answers

As Figure 1 illustrates, the recall of any response combination at another trial manifests itself as a congruence between response values. However, simple congruence does not necessarily indicate a direct relationship between trials, and ambiguities can occur. Consider the following sequence of responses:

Trial Number	Response Attribute Values	Congruence to Later Trial n
j	C <sub>j</sub> P <sub>j</sub> O <sub>j</sub>	CO
k	C <sub>j</sub> P <sub>k</sub> O <sub>j</sub>	CPO
l	C <sub>j</sub> P <sub>k</sub> O <sub>j</sub>	CPO
m	C <sub>j</sub> P <sub>k</sub> O <sub>m</sub>	CP
n	C <sub>j</sub> P <sub>k</sub> O <sub>j</sub>	

The hypothetical engram which is tapped at trial n can apparently have come from one of four sources. It is, however, important to identify a single most likely source of repetition in order to establish the chance level of its being repeated at any one trial. To do this one must make two assumptions:

- Given two possible sources with the same amount of congruence, e.g. trial k and trial l, there is no information that the later of the two is not itself a repetition of the earlier trial. In these cases the most likely source is taken to be the earliest trial in which the response combination in question appeared.
- When one possible source gives a greater match with the trial in question than another, that is taken to be the more likely source.

Taking these points, the sequential analysis can work through individual sets of data and identify the most likely source of each response by searching all previous responses and the stimuli for the earliest combination with which the response has the greatest congruence. This can produce a complex array of relationships within responses and between responses and stimuli, as shown in Figure 1.

## Identifying Chance Levels Of Repetition

Figure 1 is typical of all subjects' response sequences in showing not merely congruences between stimuli and responses, as would be expected if the subject had any

memory at all, but also between one response and another. Clearly some of these will occur by chance, and the statistical test of the significance of this process depends upon determining the chance level.

Since the total number of stimulus attribute values is finite, it follows that the chance levels of repetition at any one trial are a function of the number of different permutations of attribute values that have already occurred in the subject's response sequence. Thus, for example, the chance level,  $Q(CO)$ , for a repetition of a combination of C and O attributes is given by:

$$Q(CO) = \frac{\text{no. of different previously occurring trial combinations of C \& O}}{\text{no. of possible different combinations of C \& O values}}$$

Since the denominator is constant,  $Q(CO)$  increases gradually throughout the response sequence, as might be expected. Similar calculations can be made for other types of repetition.

Summarising the previous section, the sequential analysis can be seen as a chronological scan of response sequences from individual subjects. At each response the probability of each kind of repetition can be calculated, and the observed number of repetitions counted. It can be shown that over the entire response set the statistical significance of each type of repetition  $k$  (where  $k = CPO, CP, CO$  or  $PO$ ) over the 27 trials is given by a  $\chi^2$  statistic by the equation:

$$\chi^2 = \frac{\left( \left| \text{OBS} - \sum_1^{27} Q(k) \right| - \frac{1}{2} \right)^2}{\sum_1^{27} (Q(k)(1 - Q(k)))}$$

Where OBS is the observed number of repetitions of that type and incorporating the correction for continuity. As a result, the sequential analysis gives a value of  $\chi^2$  for each subject and for each type of answer repetitions.

The result of this is a clear and very strong trend for the repetition of complete combinations ( $\chi^2 = 1236.5, 27 \text{ df } p < .0001$ ), with all but one subject showing significant levels of this type of repetition. Less clear is the position for repetitions of smaller combinations. A small number of subjects seem to show significant frequencies of CO-repetition, although the total value of  $\chi^2$  is very much less than that for CPO-repetition ( $\chi^2 = 51.6, 27 \text{ df } p < .01$ ). There is also a *suppression* of repetitions of CP combinations by some subjects (overall  $\chi^2 = 49.1, 27 \text{ df } p < .01$ ). In view of the confused and minor role of CO-repetitions, the remainder of the analysis concentrates upon the highly significant effect of the repetition of entire CPO combinations.

## TIME COURSE OF THE REPETITION EFFECT

Figure 1 shows at trials 2 and 3 a repetition of one trial on the next. Casual observation of many such response sequences shows this to be a very frequent occurrence, and it is an interesting question whether the repetition of answers only occurs between close or adjacent trials as a short-term effect.

One property of repetitions of complete answers is that they cannot occur unless one of the components of the original response combination subsequently appears as a cue. Put another way, any occurrence of the attribute values of a response combination appearing later as cues can be taken as an *opportunity* to reproduce that combination. By the semi-random nature of the subject's responses, these opportunities will fall at different separations (in trials) from the original response. For example, in Figure 1, trial 1 could be repeated 6 and again 8 trials later, whilst trial 2 is repeatable one trial later. Of all the opportunities for repetition at any separation, a certain proportion will be successful, and the ratio of successful repetitions to opportunities gives an estimate of the probability of repetition at that value of separation.

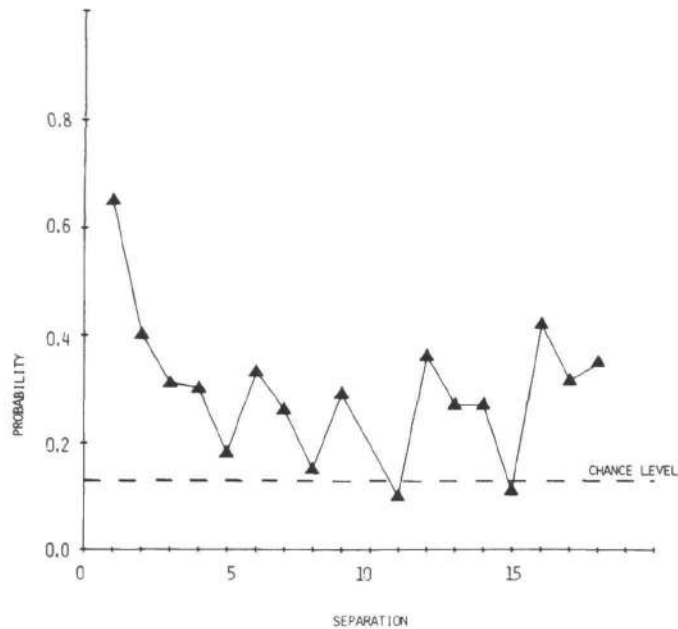
A temporal analysis was carried out in the following way. Every response combination that was not itself a completely correct response (in which case repetitions are indistinguishable from recalls of the original memory) was taken as a potential source of repetition. Each subsequent trial where one of its components appeared as cue was counted as an opportunity to repeat that combination, the intervening interval being measured in terms of the number of trials separating the two trials. In calculating the proportions of trials on which a previous response combination was repeated, the distinction was made between responses in which there was no underlying fragment and these in which a CP, CO or PO fragment existed. With only one guess included in the latter class of response combinations, the chance probability of subsequent matchings is an order of magnitude higher than in the no-fragment condition, where two guesses have to be made. For brevity, the results of CP, CO and PO fragment responses are presented together and subsequently referred to as S1 repetitions, whilst the no-fragment responses are referred to as S2 repetitions. The proportions of repetitions are plotted in Figures 2 and 3 respectively.

Both plots show high probabilities of repetition at low separations, decreasing rapidly with increasing separation to some stable level. Given that the average time between trials was some 20 to 30 seconds, this result is comparable in timescale to the long-term recency effects reviewed by Baddeley (1976, p 181).

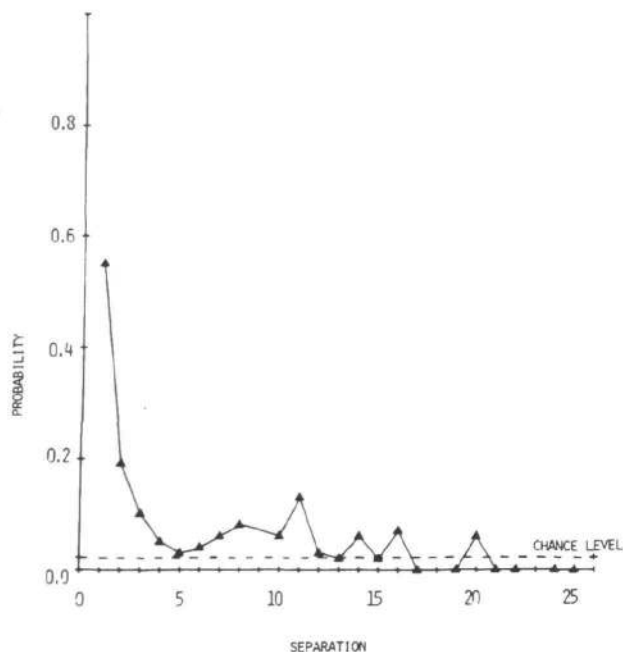
Another interesting aspect of this data is that whilst the S2 plot appears to decay to a chance level, that for S1

repetitions remains significantly above chance. Taking the data from separation 5 and above as representative of the stable portions of the curves, this can be confirmed statistically:  $\chi^2$  for S1 = 28.19, (13 df  $p < 0.25$ ), while  $\chi^2$  for S2 = 15.28, (18 df n.s.). This indicates that the mnemonic representation of S2 answers has become unavailable after a small number of trials, while some of the S1 answers remain effective for the duration of the experimental session.

**Figure 2: Probability of recalling S1 response combinations as a function of the number of intervening trials**



**Figure 3: Probability of repeating S2 response combinations as a function of the number of intervening trials**



## CONFIDENCE JUDGEMENTS

In this experiment, subjects were asked to give, in addition to their responses, a confidence rating for each attribute value given as a response. These ratings were requested in one of five categories, from 1 "SURE" to 5 = "RANDOM GUESS".

The question at issue here is whether subjects could tell the difference between genuine recalls of stimuli and recalls of previous (erroneous) answers. The ability to do so implies that memory for previous answers are additional to existing memory and distinguishable from it. A sensitive test is to look at the confidence levels of successful repetitions of S1 responses, which contain one correct recall and one guess, in comparison to non-repeated S1 responses. There are two useful comparisons here: (a) between the original and repeated S1 combinations, and (b) between the correct and error elements of a repeated S1 combination. The confidence levels used for the S1 responses were therefore broken down in three ways:

- 1) whether the response in question was correct or a guess.
- 2) whether the trial combination was a repetition of a previous trial or not
- 3) in the case of successful repetitions, whether the separation between the trials is greater than 4 trials or not, making a comparison between long term and short term repetitions which may possibly have different properties.

Within original response combinations there is, as one might expect, a significant difference in confidence for correct responses (average 1.96) as opposed to guessed responses (average 3.48), with  $\chi^2 = 40.3$ , 4 df  $p < 0.001$ . Comparison of the guesses in the repeated trials with the original guesses shows no significant differences in confidence levels in either long or short term repetitions ( $\chi^2 = 9.83$ , 8 df). Neither do the comparisons of correct responses show any differences ( $\chi^2 = 2.32$ , 8 df). It can therefore be concluded that when a subject repeats a previous S1 combination, he knows which element has been correctly recalled and which is a repetition of a previous guess. The memory of the original response combination must therefore be additional to that of the stimulus.

## DISCUSSION

To summarize the results of this analysis: in a sequence of memory trials subjects have available memory traces of previous answers which are commonly used as responses later in the sequence. These traces normally remain available only for a small number of trials, representing a duration of up to 100 seconds, but some appear to survive

for much longer periods of time. Subjects seem to be aware of the distinction between recall of these traces and recall of the original stimuli.

This analysis has a clear methodological implication: successive tests of the same memory, particularly over short intervals, will not be independent of one another. Experiments in which the observations rely upon comparison of recall on several different trials must therefore take this into account. Numerous examples can be given where doubt arises as to the validity of the conclusions in the light of this result.

What is particularly compelling about this effect is its strength and replicability across subjects, (given that no experimental instructions were made to repeat answers), and its timecourse. This corresponds well with previous work in which long term recency effects were an explicit object of study. Such effects seem difficult to explain by reference to fixed capacity stores (Waugh and Norman, 1965), or differential encoding methods (Craig and Lockhart 1972), particularly in the latter case in the light of the incidental nature of recall in this case, in which case one would not expect subjects to adopt specific strategies of encoding.

A plausible model of these long term recency effects has been proposed by Hitch et al(1980). In their view, recency effects can be explained by a strategy of retrieval based upon temporal discrimination. Items can be retrieved from memory by their temporal position in the past only while it is discriminable from the temporal position of other, related items. Long term recency effects can therefore occur when, as in these experiments, the to-be-recalled items are spaced relatively widely in time and are therefore discriminable for longer periods.

Given that the subjects were required to guess, and that memory traces for previous appropriate guesses were available, it is not particularly surprising that subjects should use them. Speculation as to why subjects should choose to repeat previous guesses is therefore not worthwhile. Whatever the subject's reasons for repeating responses, in the process of doing so he deprives the experimenter of a certain amount of information, and potentially adds some misinformation about the state of his memory. The import of this research is chiefly a methodological one: it reinforces the long held view that the experimental psychologist should be aware that the task the subjects carry out may not be quite the one he intended.

## References

- Baddeley, A.D. (1976). *The Psychology Of Memory*. Harper
- Craig, F.I.M. and Lockhart, R.S. (1972). Levels of processing: a framework for memory research. *J. Verb. Learn. Verb. Behaviour* 11:671-684.
- Hitch, G., Rejman, m. and Turner, N. (1980). *Human Memory: A new perspective on the recency effect*. Paper given at the Summer meeting of the Experimental Psychology Society, U.K. 1980.
- Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *J.E.P.* 83 (2)340-344.
- Jones, G.V. (1974). *Fragmentation of Human Memory*. Unpublished Ph.D. dissertation, University of Cambridge.
- Jones, G.V. (1976). A fragmentation hypothesis of memory: Cued recall of pictures and of sequential position. *J.E.P. Gen.* 105, 277-293.
- Kay, H. (1955). Learning and retaining verbal material. *Brit. J. Psych.* 46, 81-100.
- Loftus, E.F., Miller, D.G. and Burns, H.J. (1978). Semantic integration of verbal material into a visual memory. *J.E.P. (Human Learning and memory)* 4, 19-31.
- Lansdale, M.W. (1979). *An analysis of errors and latencies in cued recall*. Unpublished Ph.D. dissertation, University of Cambridge.
- Waugh, N.C. and Norman, D.A. (1965). Primary Memory. *Psych. Review* 72:89-104.

\* This work was carried out when the author was a research fellow at the Dept. of Psychology, University of Cambridge, U.K.





# The Role of Grid Schemata in Memory for Large-scale Environments

Ian Moar, Nancy J. Hamer, and Betsy A. Woods

Bucknell University, Lewisburg, PA 17837

Although large-scale urban environments are complex and rich in information, most of us successfully navigate such environments every day. In order to perform such a task, we must have memory representations for such environments. The present study examines how we remember large-scale urban environments.

Memory research suggests that we use higher-order knowledge structures, called schemata, to remember rich sources of information, such as stories (Thorndyke, 1977), people (Snyder and Uranowitz, 1978), rooms (Brewer and Treyens, 1981) and the complex skills involved in chess playing (Chase and Simon, 1973). Schemata may also be used to remember large-scale urban environments. Most environments of this kind comprise a grid pattern of rectangular or square shaped blocks formed by intersecting streets. A useful way to remember such areas would be in terms of a grid schema (Kuipers, 1978; Moar and Carleton, 1982). However, most urban areas do not fit perfectly into a grid pattern. For example, unlike the lines in a grid, streets are not always at right angles or parallel to each other. If we do use a grid schema to remember an urban area, such a schema may cause features of the area to be distorted in memory. In fact, we can make predictions of the kinds of distortions which may occur if we do remember urban areas using grid schemata. Five such predictions are listed below and are referred to collectively as the grid hypotheses.

1. Right-angles hypothesis. Lines in a grid always intersect at right angles. The hypothesis predicts that intersecting streets that do not meet at right angles will be falsely remembered as right angled intersections.

2. Parallel-lines hypothesis. Lines in a grid are either at right angles or parallel to each other. According to the hypothesis, if streets in an urban area travel in roughly the same direction but are not parallel, they will be misremembered as being parallel.

3. Parallel-subsets hypothesis. A single square in a grid is contained within a larger square made up of several single squares. The sides of the single square are parallel to the respective sides of the larger square in which it occurs. For example, in a 3 X 3 square grid, the sides of the center square are parallel to the respective sides of the larger square formed by all nine squares. The hypothesis predicts that if the sides of a building in a block are not parallel to the respective streets forming the block, the sides of the building will be falsely remembered as being parallel to the respective streets of the block. Here the building comprises a single square of a grid contained within a larger square, the block.

4. Straight-lines hypothesis. All lines in a grid are straight. It is predicted that streets containing bends will be remembered as being straight.

5. Alignment hypothesis. Each single square in a grid is perfectly aligned on the grid's vertical or horizontal axis with each of the four other single squares with which it shares a side. In an urban area, buildings sometimes face each other but are not perfectly aligned. Thus the middle of one building may not directly face the middle of the other building. The hypothesis predicts that such buildings will be falsely remembered as being directly aligned. In other words, the buildings will be misremembered so that the middle of one building directly faces the middle of the other building.

The five grid hypotheses were tested in three experiments. In the first experiment, students of Bucknell University drew from memory a map of part of the campus. In Experiment 2, students of the university drew a map of the whole campus from memory. Preselected features of the campus were examined on students' drawn maps to determine if the features were distorted in the manner predicted by the grid hypotheses. In both experiments, significant distortions were found for all five hypotheses, supporting the use of grid schemata in memory for the campus. In addition, half the students in Experiment 2 were freshmen and half were seniors. The seniors' drawn maps were significantly more elaborated, in terms of the number of buildings, streets or place names, than those of the freshmen. However, no significant difference was found between the freshmen and seniors in terms of the degree to which their drawn maps were distorted towards a grid pattern. The

results suggest that amount of experience with an urban area does not change the way it is remembered in terms of a grid schema.

The third experiment involved a recognition task. Students of Bucknell University were presented with maps, each of a single feature of the Bucknell campus (eg., an intersection, two streets, etc.). Nine alternative maps of each feature were presented; one correct map and eight with various degrees of distortion. For each feature, students had to choose the map which they considered was the most accurate. For each of the five grid hypotheses, students showed a significant tendency to choose the map in which the feature was distorted in the manner predicted by the appropriate grid hypothesis. Therefore, the recall results of Experiments 1 and 2 and the recognition results of Experiment 3 support the use of grid schemata in memory for the campus.

#### References

- Brewer, W.F., & Treyens, J.C. Role of schemata in memory for places. Cognitive Psychology, 1981, 13, 207-230.
- Chase, W.G., & Simon, H.A. Perception in chess. Cognitive Psychology, 1973, 4, 55-81.
- Kuipers, B. Modelling spatial knowledge. Cognitive Science, 1978, 2, 129-153.
- Moar, I., & Carleton, L.R. Memory for routes. Quarterly Journal of Experimental Psychology, 1982, 34, 381-394.
- Snyder, M., & Uranowitz, S.W. Reconstructing the past: Some cognitive consequences of person perception. Journal of Personality and Social Psychology, 1978, 36, 941-950.
- Thorndyke, P.W. Cognitive structures in comprehension and memory of narrative discourse. Cognitive Psychology, 1977, 9, 77-110.



# THE ROLES OF INTERFERENCE AND INFERENCE IN THE RETRIEVAL OF AUTOBIOGRAPHICAL MEMORIES

Brian J. Reiser and John B. Black  
Yale University

## 1. Introduction

Recent work on autobiographical memory has suggested that the representation of an individual experience is encoded within a memory structure containing generalizations about that class of events (Kolodner, 1980; Reiser, Black, & Abelson, 1982; Schank, 1982). In this paper, we examine how retrieval from such a memory category is affected by the number of events encoded within it. We argue that retrieval of personal experiences is a reconstructive process, guided by inference mechanisms that predict features of the target event for utilization as retrieval cues. More frequent events may be easier to retrieve from a category, since the circumstances leading to such an event are easy to predict. Alternatively, one interpretation of interference models of memory suggests that a greater number of events within a category would slow retrieval of one of these events. The effects of frequency of experiences are examined in an autobiographical memory experiment where subjects are asked to recall past experiences in response to verbal cues.

## 2. Inference in Memory Retrieval

Reconstructive models of memory retrieval have stressed the recursive nature of the search process: search retrieves information useful in constructing new cues for further search, until the target information is retrieved (Kolodner, 1980; Norman & Bobrow, 1979; Williams & Hollan, 1981). In particular, retrieval of an individual personal experience requires search through a memory store containing an enormous number of events. Clearly the success of such a search depends on the selection of the proper search context to restrict the portion of the data base that must be examined. This search context is selected by inference mechanisms which process the generic information in the concepts activated in the query to predict the most plausible location in memory where the target information might reside (Kolodner, 1980). Retrieval of an individual experience from memory is thus an active strategic process, directed by these inference mechanisms.

In recent reformulations of Bartlett's (1932) schema plus correction model, individual experiences are encoded in memory as associations of generic memory structures, indexed by the features on which the event deviated from the generalizations represented for that class of events (Kolodner, 1980; Schank, 1982). Search for an individual experience proceeds by using the generic knowledge about events to infer candidate memory structures for search, and then to generate indices within the structure that specify an individual experience:

*Think of a time when you felt impatient.*

Felt impatient...I'm trying to think of times when I felt impatient. Impatient always seems to mean when you're waiting in line for something or waiting for something to happen. Um...it's hard for me to think of right now for some reason. I can think of times when I felt frustrated waiting but not really impatient, and I think there's a difference. I remember waiting for someone who didn't show up for 4 hours, and it wasn't really that I was impatient, I was just frustrated with the fact that I didn't know whether this person was going to show up or not.

*Was this waiting in line?*

No, this was waiting to meet someone in front of a museum in Hartford. And he didn't come for about 4 hours and mostly the reason I was frustrated or maybe it even is impatient is that I didn't know whether he was actually coming...

This protocol demonstrates the use of inference to establish a search context. Here, the subject uses her knowledge about impatience to infer a set of circumstances in which she was likely to have experienced that affect. This process leads to thinking about situations involving "waiting," and ultimately to an experience where she was waiting for someone who was very late. Interestingly, the subjects' hesitation in classifying this experience as "impatience" supports the claim that it was accessed in memory using "waiting" as a search context, rather than simply using "impatience" as a cue.

Essentially, these type of inference processes reformulate a memory query into a set of circumstances in which the target experience might have occurred. The importance of inference in strategic memory retrieval has been



suggested by studies of retrieval protocols (Williams & Hollan, 1981) and computer models (Kolodner, 1980). A recent examination of retrieval times of autobiographical memories may be interpreted as processing evidence for the role of inference in retrieval (Reiser, Black, & Abelson, 1982). In these studies, subjects took longer to recall an experience with an activity (*took a ride on a train*) that contained a goal failure (*couldn't find a seat*) than an experience with the same activity that contained a normative action (*paid at the ticket booth*). We argued that retrieval of an experience containing a normative action typically involves less inferencing, since virtually any of the experiences stored with the activity will contain that action. Retrieval of an experience satisfying a goal failure cue requires more inferencing to find the indices that specify an experience containing the targeted type of goal failure.

In general, the difficulty of a memory retrieval is determined by the type of processing necessary to construct an appropriate search context, and to generate the indices within that context that specify an experience possessing the target attributes. Search within a memory category for an experience satisfying a particular constraint is easier if there are more types of situations in which that type of event occurs. Thus, events that occur more frequently should be easier to find within a category than less frequent events.

### 3. Interference in Memory Search

Alternatively, one might argue that more unusual experiences will be easier to find in memory, since many similar events may become confused with one another, and interfere during the retrieval process. Anderson has shown that a greater number of associations involving a concept in a memory network slows the retrieval of propositions involving that concept (e.g., Anderson, 1976). Anderson argued that the search is slowed with more associations since the limited capacity process would “fan out” over more paths. Anderson and his colleagues have observed interference in retrieval situations requiring the retrieval of a particular proposition, and have not investigated in situations where any instance of a memory category may be produced as a response. However, one possible application of interference models to this type of production task would predict that retrieval of an instance will be easier for smaller categories than larger ones, which possess more links from the category node to associated subsets and instances. Again, spreading the search over more links reduces the speed of the processing. Furthermore, if the instances possessing a particular attribute form a subcategory of the main category, then a larger number of such instances should slow the search for a category member possessing that attribute.

Although an interference model cannot account for the difference between failure and normative actions found by Reiser et al., it may be argued that this difference is not due to the frequency of the two types of events, but instead to some intrinsic difficulty in recalling goal failure events. In fact, we found retrieval time differences which one might argue support an interference model. Activities (*Going to Movies*, *Dining at Restaurants*) were found to be better retrieval cues than general actions (*paying*, *ordering*), which encode abstracted generalizations about an action (or scene) which serves as a component of several activities. We argued that the type of generic information encoded in activities is of greater utility in constructing the right combination of features to specify a unique experience. Retrieval of an experience when presented with a general action such as *waited for your turn* involves first figuring out *where* (i.e., in what kind of activity) one might have waited for something. When a candidate activity such as *Bank* or *Grocery-Shopping* has been inferred, the context-specific knowledge associated with the activity (i.e., features specific to *Bank* experiences) can be used to refine the search context.

An interference model provides an alternative explanation for this result. One might argue that the general actions are poor retrieval cues since they specify more frequent experiences than the activities (most general actions occur in many activities). Search in the category named by the general action must spread over many more links to experiences than search within an activity category, and would therefore take longer to retrieve any of the experiences.

In order to explore the relationship between the number of experiences of a given type and the difficulty of retrieving an experience from memory, we asked subjects to recall events which differed in their average frequency of occurrence. There were two motivations in the present study. First, we hoped to find that more frequent experiences within an activity would be easier to retrieve, to support the importance of inferring a search context and generating indices within that context. By using low frequency events that were not goal failures, we avoid other characteristics of goal failures that may be contributing to their retrieval difficulty. Second, presenting a case where more frequent experiences are *easier* to retrieve provides an argument against interference counter-explanations of our previous finding of retrieval differences between activities and general actions.



#### 4. Retrieval of High and Low Frequency Experiences

Norms collected by Galambos (in press) were used to construct probes about common events which varied in their frequency of occurrence. Each probe consisted of an everyday activity (e.g., *Checked Out Books*) paired with one of its component actions (e.g., *got call numbers*). Galambos' norms enable a test of the effects of event frequency at two levels. The *standardness* rating of a component action measures the within-category frequency of the action -- i.e., how likely one is to perform that action in an execution of the given activity. It is thus an indication of the relative number of experiences encoded within the activity category that involve that particular action. The activity frequency measures the size of the category itself -- a measure of how often one engages in the activity may be considered a rough estimate of the number of experiences stored in memory involving that activity.

Retrieval of experiences for trials involving highly standard actions is predicted to be faster, since it will be easier to infer a context in which this action would have been performed in the activity. For example, most *Going to Movies* experiences involve the action *gave usher tickets*, but fewer involve the action *stood in line*. Virtually any retrieval cue within the category *Going to Movies* would lead to an experience matching the first action. However, more careful consideration of the generalizations about movie-going would be necessary to predict a situation that would involve standing in line. This might lead retrieval to consider "very popular movies", "movies I saw on opening night", "movies I saw on trips to New York", etc. This retrieval thus requires more inferencing to construct a search context. This prediction is contrasted with the interpretation of the interference model suggested above, in which more frequent actions within an activity would be more difficult to retrieve.

The predictions for the effects of activity frequency are less straight-forward. If the generic information encoded in the activity is the principal source of material for inferring the search context, then the frequency of the activity itself may have a smaller effect (or none at all) on the ease of retrieving an experience. That is, if the principal component of retrieval is the inference of indices within the category, then the absolute size of the category may have little effect, and all activities may be equally accessible in memory. However, one might also expect that more frequently executed activities become more richly articulated, and thus the indices within the category may be more easily traversed. For example, recalling a time involving *Eating at a Restaurant* may require less inferencing than a more infrequent event such as *Taking a Photograph*. One is probably familiar with many contexts which include eating at restaurants (*dating, business lunches, grabbing a quick bite to save time, etc.*) while more processing might be necessary to infer a situation that would have included the taking of photographs. This would predict an advantage for more frequent activities. In either case, this prediction can be contrasted with the interference prediction of slower retrieval times for the larger categories.

**Method.** Two actions, one high and one low in standardness, were selected from each of 18 of the activities presented in Galambos (in press). The high and low standard actions were selected to be equal in "centrality" (the importance of the action to the goals of the activity), since subjects recall the more central actions of an experience more easily than the less central actions (Reiser, in preparation).

The autobiographical retrieval task developed by Reiser, Black, and Abelson (1982) was used to measure the accessibility of the target events in memory. Subjects were told that each trial would consist of two phrases: a description of an activity, and a description of an action that takes place in the activity. Upon initiation of the trial by the subject, the two phrases were simultaneously displayed on a CRT, with the activity phrase presented above the action phrase. Subjects were to recall an experience where they were performing the stated action while doing the stated activity. The subject responded by pressing a *Yes* key on recall of such an experience, or pressed the *No* key if he or she could not recall an experience. We emphasized that the memory be of a specific experience, but that it was not necessary to recall all of the details before responding. Subjects wrote a one or two sentence description of the experience immediately following each *Yes* response. At the completion of the reaction time task, subjects were asked to provide the month and year in which each of their recalled experiences had occurred. Each subject was probed about all 18 activities, but to avoid possible priming effects, a subject was probed only once about each activity. Each subject thus received 9 activity-action pairs containing a high standard action, and 9 activity-action pairs containing a low standard action. Forty-eight Yale undergraduates participated in this experiment for course credit.

**Results.** Of primary interest are the retrieval times for those trials in which subjects successfully recalled an experience. The mean retrieval times and proportion recalled for *Yes* responses are:

Activity + High Standard Action	2.929 sec; 84% <i>Yes</i> responses.
Activity + Low Standard Action	3.359 sec; 77% <i>Yes</i> responses.

As predicted, subjects were faster to recall an experience involving a high standard action [ $\min F(1,64) = 4.19, p < .05$ ]. This supports the prediction that more processing is required to retrieve an experience involving a less frequent component action.

In order to assess the separate contributions of action standardness and activity frequency, we performed a multiple regression analysis of the mean retrieval times for each activity and action pair. The activity frequency did not significantly affect retrieval time [ $F < 1$ ]. Thus, retrieval from large event categories (frequently executed activities) was no more difficult than from smaller categories.

One possibility is that the high standard actions were easier to retrieve because they had been performed in a more recent experience with the activity. In fact, high standard actions tended to elicit more recent experiences than low standard actions (4.1 vs. 5.5 months), but this difference was not significant ( $p > .10$ ). More importantly, when age of experience is used as covariate in an analysis of covariance of retrieval times, the difference between high and low standard actions remains significant. Finally, the number of experiences recalled by subjects decreased with the elapsed time since the experience, as expected from previous studies (e.g., Crovitz & Schiffman, 1974). Interestingly, a median split on retrieval times for each condition revealed that older experiences were generally slower to retrieve [ $\min F(1,54) = 3.63, .05 < p < .10$ ].

**Discussion.** This experiment has demonstrated that retrieval of a more frequent experience within a given activity is easier than that of a less frequent experience. This is consistent with a reconstructive model of autobiographical retrieval. Processing first retrieves the category, accessing the generic knowledge encoded about these events. This generic knowledge is utilized to construct retrieval cues consisting of the circumstances likely to have resulted in the target type of experience. Searching for an instance of a frequently executed action is relatively easy, because virtually any context within that activity would involve experiences including the target action. Retrieval of a less frequent event requires greater use of the generic information to infer the *type* of experience within the activity that would include the targeted action.

The failure of category size (i.e., activity frequency) to affect retrieval times suggests that smaller categories are no easier to access than larger categories. The prediction from interference theories of slower retrieval times for larger categories was not supported. Instead, the results support the suggestion that the principal inferencing in retrieval involves constructing the proper path to the event within the category, and that the categories themselves do not differ in their accessibility.

The slower retrieval times for the older experiences conflicts with the curvilinear pattern found by Robinson (1976), where events from 0-5 years and 10-15 years were recalled more quickly than events from 5-10 years ago. Since the data from the present experiment are almost exclusively from the 0-5 year interval, we can not evaluate them for Robinson's curvilinear pattern, and can conclude only tentatively that the older experiences recalled by subjects were less accessible in memory. In fact, the slower retrieval times for the older experiences may be explained by a retrieval strategy that focuses on the subject's current context for retrieval cues. For example, suppose the subject uses knowledge about his or her current job, place of residence, school, social situation, etc. to generate cues for recall. The cues generated in this fashion would be more consistent with recent experiences than with older experiences. If retrieval fails to recover an experience, older contexts may then be tried. Thus, the subject recalls fewer older experiences, and these trials are slower than those retrieving more recent experiences.

Finally, these results argue against an interference explanation of the retrieval differences between activities and general actions. There is no evidence to support slower retrieval times for larger categories; in fact, we have shown that in some situations, more frequent events are easier to retrieve. In combination, these results suggest that structural features such as "frequency" may be inadequate to fully explain memory search. When comparing activities and general actions, more frequent events are *more difficult* to retrieve; when comparing the retrieval of two actions within an activity that differ in frequency, more frequent events are *easier* to retrieve. Clearly, the difficulty of search from a concept is a function not only of the number of associations involving that concept, but also of the type of concept and the type of information encoded within such a category (Reiser & Black, 1982).

### Acknowledgments

Bob Abelson, Jim Galambos, and John Anderson provided valuable comments on an earlier draft of this paper. We also thank Maren Jones for assistance in the collection and analysis of the data. This research was supported by a grant from the System Development Foundation.

## References

- Anderson, J.R. *Language, memory, and thought*. Hillsdale, NJ: Erlbaum, 1976.
- Bartlett, F.C. *Remembering: A study in experimental and social psychology*. New York: Cambridge University Press, 1932.
- Crovitz, H.F., & Schiffman, H. Frequency of episodic memories as a function of their age. *Bulletin of the Psychonomic Society*, 1974, **4**, 517-518.
- Galambos, J.A. Normative studies of six characteristics of our knowledge of common activities. *Behavior Research Methods and Instrumentation*, in press.
- Kolodner, J.L. *Retrieval and organizational strategies in conceptual memory: A computer model*. Technical Report #187, Department of Computer Science, Yale University, 1980.
- Norman, D.A., & Bobrow, D.G. Descriptions: An intermediate stage in memory retrieval. *Cognitive Psychology*, 1979, **11**, 107-123.
- Reiser, B.J. *Contexts and indices in autobiographical memory*. PhD dissertation (in preparation), Cognitive Science Program, Yale University, 1983.
- Reiser, B.J., & Black, J.B. Processing and structural models of comprehension. *Text*, 1982, **2**, 225-252.
- Reiser, B.J., Black, J.B., & Abelson, R.P. Retrieving memories of personal experiences. *Proceedings of the Fourth Annual Conference of the Cognitive Science Society*, Ann Arbor, MI, 1982.
- Robinson, J.A. Sampling autobiographical memory. *Cognitive Psychology*, 1976, **8**, 578-595.
- Schank, R.C. *Dynamic memory: A theory of reminding and learning in computers and people*. New York: Cambridge University Press, 1982.
- Williams, M.D., & Hollan, J.D. The process of retrieval from very-long term memory. *Cognitive Science*, 1981, **5**, 87-119.



# **Paper Session #9**





# Thematic Relations Between Episodes

Colleen M. Seifert and John B. Black  
Yale University

Among the variety of knowledge structures proposed to capture information in episodes, a dichotomy between contentful structures and content-free structures appears useful in characterizing abstract relations in memory (Schank, 1982). The level of relatively content-free structures is needed to explain first, the thematic pattern within an episode, and second, how generalizations are made across episodes that vary greatly in some respects while sharing more abstract similarities.

Plot Units (Lehnert, 1981) have been proposed to capture knowledge about thematic concepts that represent particular patterns of goal relationships and events. Plot units such as *competition*, *failure*, and *shared positive events* can be used to describe the goal situation in an episode. Their utility in understanding, generating, and summarizing narratives has been demonstrated (Reiser, Black, and Lehnert, 1982; Lehnert, Reiser, and Black, 1981).

Although plot units involve relatively abstract knowledge about patterns of goals, they appear to be at a low level of thematic information. Higher level patterns of thematic information seem to exist which provide interesting generalizations that plot units do not capture. These thematic patterns serve to capture a level of information that is useful in building memory structures; that is, to store episodes, thereby providing connections to related experiences. Plot units appear to play a role as components in the construction and recognition of these higher level thematic knowledge structures. Information at the plot unit level is needed to access appropriate thematic structures in memory. In fact, some combinations of plot-unit-like components are likely to be higher level thematic structures.

Let's examine an example of plot unit components to consider their possible incorporation into memory structures. The *failure* plot unit is the basis for countless episodes. Combining the *failure* plot unit with *shared negative event* specifies more of a plot, but still is a rather general experience. For example, this story is used in Reiser, Black and Lehnert (1982) as an exemplar of this pattern:

Kennedy had struggled admirably to stay alive in the primaries, but his efforts were not able to win him the nomination. When his delegates finally accepted the inevitable defeat, there were more than a few tears and hollow hopes for 1984.

If this story is extended in a certain way, we introduce a theme which plot units do not capture. For example, this story could continue on to describe Kennedy's successful nomination in 1984. We would have a different thematic pattern, captured by the adage, *if at first you don't succeed, try, try again*, as well as a more interesting and memorable story. The thematic information captured by this adage is not expressed by the plot units *failure*, *shared negative event*, and *success*. The adage captures a pattern of goals and plans in which plot units appear as pieces. If the pattern is made more distinctive, such as trying over and over and never succeeding, adding the context of elections would probably remind some people of Adlai Stevenson. In this way, pieces of plot unit-like information are combined into specific patterns that not only appear to store episodes, but can produce reminders in appropriate contexts.

Here is another example of how a small change in an episode can change the meaning and result in a more distinctive, complex structure:

### Minister's Complaint

In a lengthy interview, Reverend X severely criticized the then 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 *The Church Today* magazine.

This is another routine news report of a clergyman complaining about pornography. However, if a small change is made to the story -- if *The Church Today* magazine is changed to *Penthouse*-- the point or moral of the story is drastically different. When reading the changed story, the theme may bring other experiences or even adages to mind, such as: *The pot calling the kettle black*, *Practice what you preach*, and *Throwing stones when you live in a glass house*.

How can we determine which patterns of goal and plan situations are distinctive and useful in memory organization? Schank (1982) has suggested examining the similarities and common themes in everyday life that are revealed when one is reminded of thematically similar episodes. The patterns of goal and plan interactions that people recognize and share, and that remind people of other experiences, are sometimes captured in common sayings and reminders. Adages often serve as an effective way of characterizing the theme of an individual episode. It would seem that cultural sayings, based on the commonalities of experiences in a variety of settings, are a good place to begin looking for structures to organize episodes in memory.

Dyer (1982) has used this approach to develop Thematic Abstraction Units (TAUs) which seem to capture the thematic level of adages that is not captured by plot units. TAUs are based on the analysis of adages, in particular ones that represent expectation failures that occur due to errors in planning. A TAU contains an abstracted planning structure that tells where the error was, and can serve as a warning for the planner. They serve as episodic memory structures which organize events which involve similar kinds of planning failures.

For example, here is a TAU structure for the Minister's Complaint story (Dyer, 1982):

#### TAU-HYPOCRISY

X is counter-planning against Y  
 X is trying to get a higher authority Z to block or punish Y  
     for using Plan P1 by claiming P1 is unethical  
 X has also used the unethical plan P1  
 therefore, X's strategy fails.

In the revised Minister's Complaint story, the minister tries to move public opinion against Carter by claiming Carter supports pornography. Since the Minister supported it to the same degree, his strategy fails.

In this manner, TAUs attempt to capture the goal and plan interactions that include particular errors in planning. These structures explain relations between elements within the story, and serve as the basis for connections between related episodes. From this point of view, the important questions for thematic knowledge structures are concerned with explaining their use in memory to store experiences, thereby providing a basis for bringing to mind related episodes needed for learning. But before we can address the use of thematic structures in the storing and retrieving of abstractly related episodes, it is appropriate to gather some evidence on whether people can recognize and use thematic similarity. Are the themes recognized as a wholistic pattern, rather than as related components? Do people perceive the thematic structure as more important than content similarities? We used the methodology in Reiser, Black and Lehnert (1982), employed to study plot units, to examine stories based upon TAU relations in the tasks of narrative generation and sorting.

## Story Generation Experiment

Asking subjects to write stories based on prototypical stories should indicate whether they are able to abstract the thematic similarities in the prototypes and reproduce that theme in a new context. In this experiment, subjects were given three example stories based on TAUs suggested by Dyer (1982) and were asked to write "one new story that has the same type of plot." The subjects were told not to use the same events from the examples. Each subject was given three sets of example stories and was asked to write three stories. Sixty-six subjects participated in the experiment, though not all the subjects were able to complete the task in the allotted time (15 minutes).

Here is an example of a subject-generated story based on prototypes such as "the Minister's Complaint" story discussed above:

Sue's basketball coach was outlining her pre-season training program. "Stay away from overeating, alcohol, and smoking. Everyone knows an athlete should respect her body," she said, as she drew on a cigarette.

Though some stories were based on contexts present in one of the nine exemplars for each subject, most stories had a context novel to the experiment; for example, crime, music, psychology experiments, and most popularly, "college life". The stories were scored by a trained coder who did not know which examples the stories were based upon. A proportion of how many of the stories had a recognizable and the correct TAU structure was determined, as shown in Table 1.

Table 1  
Story Generation Results

TAU	Associated Adage	Proportion Based on TAU	Total Number of Stories
Hypocrisy	Pot Calling the Kettle Black	.96	26
Incompetent Advice	Blind Leading the Blind	.88	26
Unsupported Plan	Counting Chickens Before Hatched	.78	27
Acting Too Late	Closing Door after Horse Gone	.89	19
Plan Backfires	Cutting Off Nose to Spite Face	.80	15
Self-Deception	Hiding Your Head in the Sand	.60	15
Too Costly	Killing Fly With Elephant Gun	.88	8
Leave Alone	Cure is Worse than Disease	.70	10
Leaderless	Too Many Cooks Spoil Broth	.88	8

An example of a subject's story that confuses two TAUs demonstrates the possible interactions of TAU structures active in an episode. This story contains the elements of both *TAU-INCOMPETENT-ADVICE* and *TAU-UNSUPPORTED-PLAN*.

Frank was at the horse races. He had heard from his friend, a jockey, of a sure win. He took \$1000 from his bank account, because he was so sure he would win a sure thing. When the race time rolled around, his horse was neck and neck, but came in second, and he lost his money.

In general, subjects were very good at generating the matching internal TAU patterns in a new setting. A more interesting test is how subjects perceive a series of episodes. The story sorting task provides an indication of the thematic similarities the subjects attend to.

## Story Clustering Experiment

Six subject-written stories were chosen randomly from six TAU groups: *TAU-HYPOCRISY*, *TAU-INCOMPETENT-ADVICE*, *TAU-UNSUPPORTED-PLAN*, *ACTING-TOO-LATE*, *TAU-PLAN-BACKFIRES*, and *TAU-SELF-DECEPTION*. The thirty-six unedited stories were presented in one of four random orders to a separate group of thirty-six subjects. Each subject was asked to sort the stories into groups with "similar plots" (Reiser, et al). Subjects were not told how many groups to form, but two to ten groups was suggested as a guideline. After completing the sorting task, subjects were asked to go back and label the groups with a descriptive phrase.

A hierarchical clustering analysis (Johnson, 1967) revealed how strongly pairs of stories were related (i.e., how often two stories were sorted into the same group by different subjects). The clustering results are shown in Figure 1.

Figure 1

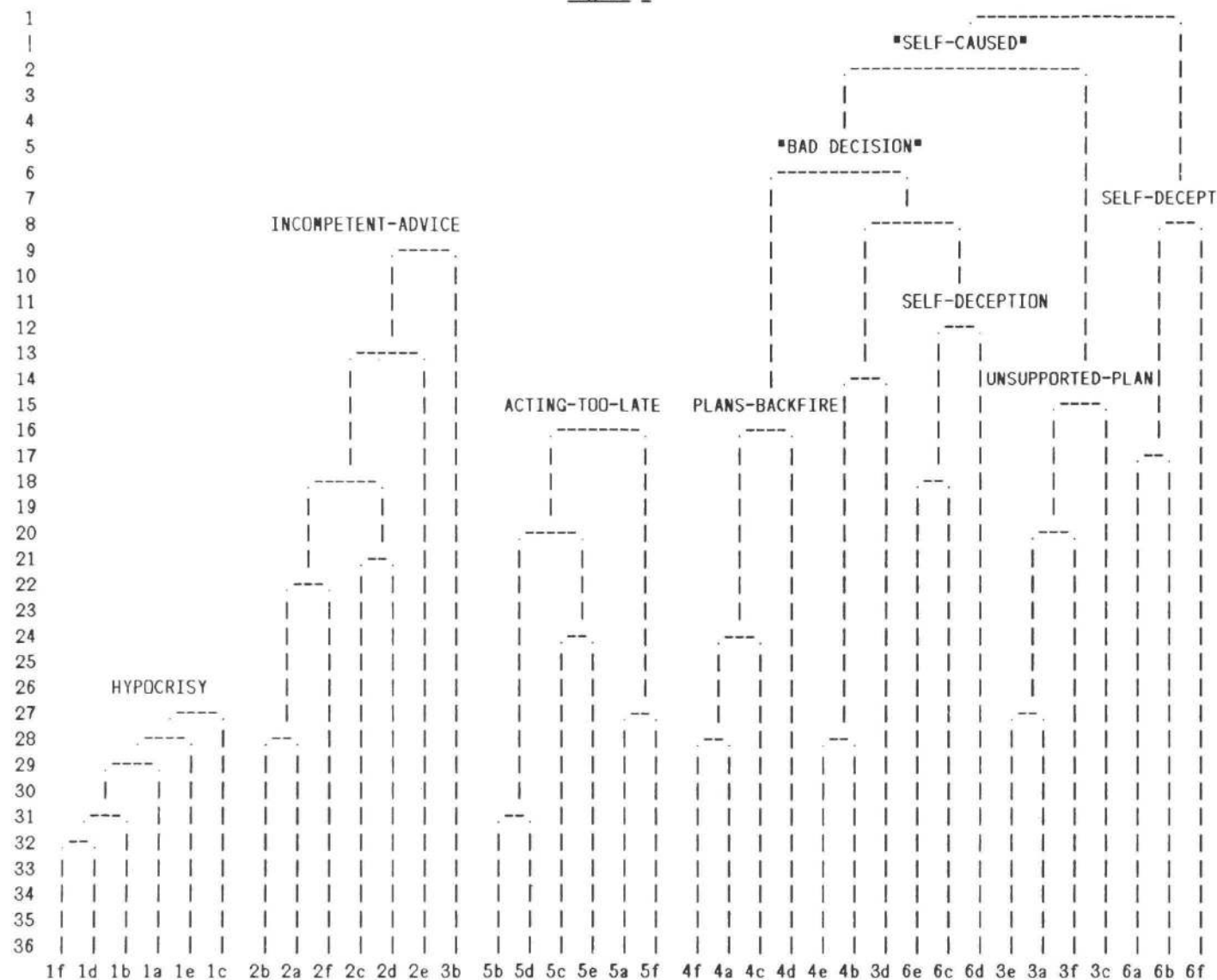


Figure 1: The hierarchical clustering diagram for the 36 subject-written stories sorted by 36 other subjects. The numbers on the horizontal axis represent similarity scores.

In general, subjects seemed to grasp the TAU structures as wholes rather than their subparts. This analysis produced three clearly defined clusters, each corresponding to a specific TAU, and a more complicated cluster involving the other three TAU structures. While two of the TAU groups are fairly well preserved in

this large cluster, one group *TAU-SELF-DECEPTION*, is split in half. A analysis of the commonalities in the labels for this group indicates subjects divided *self-deception* stories into two types: in one, the error is a bad decision which is the actor's own fault (self-caused); in the other, *self-deception* stories where the subject truly not aware they are deceiving themselves. An example of this dichotomy is a story where a student chooses to ignore her work, resulting in failure, compared to a story where a mother can't admit to herself that her son is delinquent, and he later gets in trouble. This factor of whether the actor *causes* his own disaster, or whether circumstances control his fate, is not treated systematically by TAUs. This causation factor must be incorporated into the TAU structure in order to explain the variations within tau-related episodes.

Within the TAU-based clusters, subjects tended to consider the stories more similar if they included content similarities as well. For example, stories 4b and 4e were based on *TAU-PLANS-BACKFIRE*, and were both about running away from home only to run into trouble. Other stories rated with high similarity also contained common content, such as 1d and 1f, which were both about smoking.

The analysis of the labels subjects used to describe the groups they formed revealed a high degree of agreement even in the words used. For example, 24 subjects used "hypocrisy" in their label, 24 used "too late", and 20 used "bad advice". Beyond the TAU-based labels, subjects appeared to form some groups using more general similarities, particularly "bad decision" and the "self-caused" distinction. In addition, many subjects used adages to label a group, including *practice what you preach*, *counting your chickens before they hatch*, and *the ostrich syndrome*.

## Conclusion

In summary, subjects were able to preserve the TAU pattern in their stories so that it is recognizable to other subjects, and subjects were able to use TAUs as the basis for story similarity. These experiments, as an initial undertaking, demonstrate subjects' sensitivity to thematic patterns, and indicate the thematic level of information can be used when indicated. It is clear that the thematic information present in an episode plays a crucial role. Further experimentation on the representation and the specific processing functions of the thematic knowledge will serve to determine the nature of their role in understanding.

Acknowledgements: This research was supported by grants from the Sloan Foundation and the System Development Foundation. We would like to thank Michael G. Dyer for valuable assistance during the course of this research and Steven Lytinen for comments on an earlier draft of this paper.

## References

- Dyer, M. G. In-Depth Understanding: A Computer Model of Integrated Processing for Narrative Comprehension. Research Report #219, Department of Computer Science, Yale University, 1982.
- Johnson, S. C. Hierarchical Clustering Schemes. *Psychometrika*, 1967, **32**(3).
- Lehnert, W. G. Plot Units and Narrative Summarization. *Cognitive Science*, 1981, 5, 293-331.
- Lehnert, W. G. , Black, J. B. and Reiser, B. J. *Summarizing Narratives*. Proceedings of the 7th International Joint Conference On Artificial Intelligence, Vancouver, BC, 1981.
- Reiser, B. J., Black, J. B., and Lehnert, W. G. *Thematic Knowledge Structures in the Understanding and Generation of Narratives*. Technical Report #16, Cognitive Science Program, Yale University, 1982.
- Schank, R. C. *Dynamic Memory: A theory of reminding and learning in computers and people*. New York: Cambridge University Press, 1982.





# READING A PROGRAM IS LIKE READING A STORY (WELL, ALMOST)

Elliot Soloway, Kate Ehrlich, Eric Gold  
Dept. of Computer Science  
Yale University  
New Haven, Ct. 06520

## 1. A Schemata-Based Theory of Program Comprehension<sup>1</sup>

A computer program, a mathematical proof, and an electronic circuit diagram all are (1) products of a problem solving process that required specific technical knowledge and (2) can be "executed" to obtain a specific result. However, these entities can also communicate information beyond simply the desired specific goal: in reading such "problem solving texts" and analyzing the techniques, style, comments, digressions, etc. one can gain insight into the problem itself — and even into the problem solver himself. "Stories" can also be conveyed as texts; on the other hand, they typically are not the products of technical knowledge nor are they executed for a specific result. As such, stories differ from problem solving texts. However, these two text forms are similar when both are used as a communicative vehicle. This similarity between the two text forms serves to raise an intriguing question: *Do the information representations and processing strategies that underlie the comprehension of stories also underlie the comprehension of "problem solving texts"?* In particular, can the schemata-based approach to story understanding be productively used in developing a theory of how programmers read and understand computer programs? In this brief paper, we will outline an affirmative answer to this question and describe one empirical study that supports our position. (See also [15, 16, 7, 9].)

The term we have used to express the notion of schema in the domain of computer programs is *programming plan*. Just as a schema [13, 3, 2] captures generic knowledge, a programming plan specifies the critical information that is representative of the stereotypic action sequences in programs. For example, we can identify two types of programming plans in the program in Figure 1: control flow plans and variable plans.<sup>2</sup> For example, the RUNNING TOTAL LOOP PLAN and the SKIP GUARD PLAN are two control flow plans in this program. The former plan repeatedly reads in some values and accumulates their total. The latter plan is also a common one: it protects the main computation of the loop from an illegal input value; should the value be input, the main processing steps are skipped over. Variable plans serve to highlight the role a variable plays in a program: just as actors take on different roles in a play, variables take on different functions in a program. For example, the COUNTER VARIABLE, Count, is used to count the number of elements being accumulated, e.g., `Count:=Count+1`. Similarly, the RUNNING-TOTAL VARIABLE is used to accumulate a total, e.g., `Sum:=Sum+Num`. While both variables are updated using an assignment statement, programmers do seem to distinguish between them on the basis of their functions [16].

## 2. Generating Plan-like and Unplan-like Programs

What makes a program plan-like rather than unplan-like is the way in which plans are composed in a program. In particular we have identified two rules of plan composition that can be used to systematically vary the planliness of a program. These rules are: (1) vary the typicality of the plans being composed into a program, (2) modify a typical plan in an atypical manner (usually to let the plan do "double duty"). In Figure 1, we illustrate the effect of applying these rules. The programs, which all solve the problem given in the figure, were generated using the two rules above. Programs A, B in Figure 1 reflect compositions of increasingly less typical plans. Program C reflects compositions of typical plans that have been modified in such a way as to be atypical. The heart of the problem in this figure requires a SENTINEL-CONTROLLED RUNNING TOTAL LOOP PLAN. The key issues are to add up the numbers being read in while keeping the sentinel value from being added into the total and keeping the count from also being updated. What

---

1

This work was co-sponsored in part by the Personnel and Training Research Groups, Psychological Sciences Division, Office of Naval Research and the Army Research Institute for the Behavioral and Social Sciences, under Contract No. N00014-82-K-0714, Contract Authority Identification Number, Nr 154-492. Approved for public release; distribution unlimited. Reproduction in whole or part is permitted for any purpose of the United States Government. This work was also sponsored in part by NSF RISE under grant number SED-81-12403.

<sup>2</sup>Variable plans are related to, but are richer than, the computer science notion of *abstract data types*, in that plans have more properties (e.g., relatedness, goal) than are usually associated with abstract data types.

follows is a detailed plan analysis of these programs.

- In Pascal, SENTINEL-CONTROLLED RUNNING TOTAL LOOP PLAN is most appropriately realized with a WHILE looping construct [17](Figure 1A). The sentinel is prevented from corrupting the loop in the following manner:
  - ▶ a Read of the input is positioned before the loop begins
  - ▶ if the sentinel value turns up on this first Read, it will be detected *before* the loop is executed even once; in this case processing with drop down to the IF statement.
  - ▶ if the sentinel value does not turn up on the first Read, then this legitimate value is added into the running total, Sum, and the counter, Count, is updated accordingly.
  - ▶ after these updates occur, the next value is Read in and processing returns to the top of the loop where the new value is tested; should this value be the sentinel, processing will drop down to the IF statement without further processing in the loop (i.e., without adding the sentinel into the running total).

The COUNTER and RUNNING TOTAL VARIABLES employ the standard VARIABLE PLAN initialization and update techniques: start the value off at 0, and update appropriately. Thus, the composition of the variable plans and the loop plan is accomplished using standard techniques.

- The program in Figure 1B, however, does not use a WHILE loop, but rather a REPEAT loop. In order to protect the running total and the count from being incorrectly updated, a SKIP GUARD PLAN is used that encloses these update steps. In other words, there is a *causal relationship* between the LOOP PLAN and the GUARD PLAN: we need the GUARD PLAN to make up for the LOOP PLAN's inadequacy. SKIP GUARD plans are typical techniques in programming; we see one used to protect the average calculation from a divide by 0 case. Again, the COUNTER and RUNNING TOTAL VARIABLES employ the standard VARIABLE PLAN initialization and update techniques. While the composition of the variable plans and the loop plan *and* the skip guard plan is accomplished using standard techniques, it is less typical to realize a SENTINEL CONTROLLED LOOP PLAN with a REPEAT loop composed with a SKIP GUARD PLAN. This judgement of typicality is based on experience in teaching Pascal from numerous textbooks and on articles describing good programming style [17].
- While the program in Figures 1C still achieves the overall objective, it was constructed by taking standard plans and modifying them in an atypical manner. For example, the sentinel value must again be backed out of the running total variable and the counter variable. This time, however, the initialization technique of the two variable plans are modified to serve this additional function: to say the least, initializing a variable to -99999 is a very curious construction.

Does planliness effect program comprehension? One of the most important implications of a schema is that it provides a structure for comprehending and encoding information. Researchers have shown that a story is remembered better if it is more schema-like e.g. [2, 11, 5]. Similar results have been obtained for comprehension in non-story domains [6]. In the next section we will present one study in which we examined the this the issue of planliness and program comprehension using versions of the programs shown in Figure 1.

### 3. Empirical Evidence: A Taste

Advanced programmers (end of at least second semester of programming) were split into two groups; half were presented with program Alpha in Figure 2, while the other half were presented with the program Beta. Both groups were asked to fill in the blank line with a line of code that, in their opinion, most reasonably completes the program. Subjects were not told what problem the program was intended to solve. A version of this technique was used by Bower, Black, and Turner [3] and Kemper [10] in order to tap into the schemata people used in comprehending stories. Our hypothesis is that if programmers are using programming plans to comprehend the programs, then the expectations set up by those plans will make it easier to fill-in-the-blank in the more plan-like programs (Alpha, Figure 2). However, we suggest that it will be more difficult to comprehend the less plan-like program (Beta, Figure 2), since few expectations will be set in motion.

The results are displayed in Figure 2C. The correct answer for problem Alpha was `Count := 0`, while the correct answer for Beta was `Count := -1`. Based simply on the number of correct and incorrect answers, it was clear that program Beta elicited very different performance from that of program Alpha: there were more correct responses to Program Beta than to Program Alpha ( $X^2 = 47.7$ ,  $p < 0.001$ ). Moreover, it is not just that there are differences in the accuracy of the responses but also that subjects took longer to give their responses in the unplan-like program. An analysis of variance on the time to read the program and fill in the blank reveals that subjects took longer to read the unplan-like program (Beta) than to read the plan-like program (Alpha) ( $F[1,91] = 4.60$ ,  $p < 0.05$ ). There was no difference in reading time between correct and incorrect responses ( $F[1, 91] = 3.06$ ,  $p > 0.05$ ). The result that is particularly interesting is that there is an interaction between the factors of program and response: the difference in response time between correct and incorrect responses is much greater for Program Beta than for Program Alpha ( $F[1, 91] = 4.50$ ,  $p < 0.05$ ). That is, subjects took longer to get Beta correct than to get Alpha correct. These results lend

strong support to our claim that experienced programmers use their knowledge of plans to comprehend programs and that they will therefore take longer to comprehend unplan-like programs correctly than to comprehend plan-like programs correctly.

Interestingly, standard software engineering metrics of program complexity such as (1) lines of code or (2) a Halstead [8] metric, predict that program Beta, with fewer lines, less volume, and fewer nested structures would be easier to comprehend than program Alpha. However, given our plan-based analysis, we have argued for Alpha being the less complex — and the experimental data cited above supports this position.

#### 4. Concluding Remarks

In this brief summary, we have attempted to indicate the direction in which our research into program comprehension is going. We have given a brief description of how one can create plan-like and unplan-like programs, and we have described results from one experiment in which we used these programs in order to examine the use of schemata in program comprehension. These results are consistent with, but more fine-grained than, previous work on the role of schemata in technical domains in general [6, 4, 5], and programming in particular [14, 1, 12].

1. Adelson, B. "Problem Solving and the Development of Abstract Categories in Programming Languages." *Memory and Cognition* 9 (1981), 422-433.
2. Bartlett, F.C.. *Remembering*. University Press, Cambridge, 1932.
3. Bower, G.H., Black, J.B., Turner, T. "Scripts in Memory for Text." *Cognitive Psychology* 11 (1979), 177-220.
4. Chase, W.C. and Simon, H. "Perception in Chess." *Cognitive Psychology* 4 (1973), 55-81.
5. Chiesi, H.L., Spilich, G.J. and Voss, J.F. "Acquisition of domain-related information in relation to high and low domain knowledge." *Journal of Verbal Learning and Verbal Behavior* 18 (1979), 257-273.
6. deGroot, A.D.. *Thought and Choice in Chess*. Mouton and Company, Paris, 1965.
7. Ehrlich, K., Soloway, E. An Empirical Investigation of the Tacit Plan Knowledge in Programming. in *Human Factors in Computer Systems*, J. Thomas and M.L. Schneider (Eds.), Ablex Inc., in press.
8. Halstead, M.M.. *Elements of Software Science*. Elsevier, New York, 1977.
9. Johnson, L., Draper, S., Soloway, E. Classifying Bugs is a Tricky Business. NASA Workshop on Software Engineering, in press.
10. Kemper, S. "Filling in The Missing Links." *Journal of Verbal Learning and Verbal Behavior* 21 (1982), 99-107.
11. Kintsch, W., van Dijk, T.A. "Toward a Model of Text Comprehension and Production." *Psychological Review* 85 (1978), 363-394.
12. McKeithen, K.B., Reitman, J.S., Rueter, H.H., Hirtle, S.C. "Knowledge Organization and Skill Differences in Computer Programmers." *Cognitive Psychology* 13 (1981), 307-325.
13. Schank, R.C. and Abelson, R.. *Scripts, Plans, Goals and Understanding*. Lawrence Erlbaum Associates, Hillsdale New Jersey, 1977.
14. Shneiderman, B. "Exploratory Experiments in Programmer Behavior." *International Journal of Computer and Information Sciences* 5,2 (1976), 123-143.
15. Soloway, E., Bonar, J., Ehrlich, K. . Cognitive Strategies and Looping Constructs: An Empirical Study. Communications of the ACM, in press.
16. Soloway, E., Ehrlich, K., Bonar, J., Greenspan, J. What Do Novices Know About Programming? In A. Badre, B. Shneiderman, Ed., *Directions in Human-Computer Interactions*, Ablex, Inc., 1982.
17. Wirth, N. "On the Composition of Well-Structured Programs." *ACM Computing Surveys* 6, 4 (1974).

Problem Read in numbers, taking their sum, until the number 99999 is seen Report the sum Do not include the final 99999 in the sum

(A)

```

PROGRAM OrangeAlpha.
VAR Sum, Count, Num  INTEGER.
    Average  REAL.
Counter Variable
Plan
-----> Count := 0.
|
|  --> Sum := 0.
|  Read(Num).
|  WHILE Num <> 99999 DO
|  BEGIN
|  Sum = Sum + Num.
|  Count = Count + 1.
|  Read(Num).
|  END
|  IF Count > 0 THEN
|  BEGIN
|  Average = Sum/Count.
|  WriteIn( Average).
|  END
|  ELSE
|  WriteIn( 'no legal inputs').
|  END
END

```

Running Total Loop Plan

Running Total Variable Plan

Skip Guard Plan

(B)

```

PROGRAM OrangeB.
VAR Sum, Count, Num  INTEGER.
    Average  REAL.
BEGIN
Sum = 0.
Count = 0.
REPEAT
Read(Num).
IF Num <> 99999 THEN
BEGIN
Sum = Sum + Num.
Count = Count + 1.
END.
UNTIL Num = 99999.
IF Count > 0 THEN
BEGIN
Average = Sum/Count.
WriteIn( Average).
END
ELSE
WriteIn( 'no legal inputs').
END
END

```

Running Total Controlled Running Total Loop Plan  
implemented with a REPEAT Loop Plan that  
realizes a Read-Process Loop Strategy  
composed with a Skip Guard Plan  
to simulate a Sentinel-Controlled Running Total Loop Plan  
using Count Variable Plan  
Running Total Variable Plan  
New Value Variable Plan  
Skip Guard Plan  
using average calculation

(C)

```

PROGRAM OrangeC.
VAR Sum, Count, Num  INTEGER.
    Average  REAL.
BEGIN
Sum = -99999.
Count = -1.
REPEAT
Read(Num).
Sum = Sum + Num.
Count = Count + 1.
UNTIL Num = 99999.
IF Count > 0 THEN
BEGIN
Average = Sum/Count.
WriteIn( Average).
END
ELSE
WriteIn( 'no legal inputs').
END
END

```

Running Total Controlled Running Total Loop Plan  
implemented with a REPEAT Loop Plan that  
realizes a Read-Process Loop Strategy  
composed with a Patch Plan  
Modify Initialization of Count Variable  
Modify Initialization of Running Total Variable  
to simulate a Sentinel-Controlled Running Total Loop Plan  
using Count Variable Plan  
Running Total Variable Plan  
New Value Variable Plan  
Skip Guard Plan  
using average calculation

Figure 1: Examples of Programming Plans

```

(ALPHA)

PROGRAM OrangeAlpha.
VAR Sum, Count, Num: INTEGER.
    Average: REAL.
BEGIN
    Sum = 0.
    -----
    |                                     |.
    -----
    REPEAT
        Readln(tty); Read(tty, Num);
        IF Num <= 99999 THEN
            BEGIN
                Sum = Sum + Num;
                Count = Count + 1;
            END;
        UNTIL Num = 99999;
        Average = Sum/Count;
        Writeln(tty, Average);
    END
END

```

```

(BETA)

PROGRAM OrangeBeta.
VAR Sum, Count, Num: INTEGER.
    Average: REAL.
BEGIN
    Sum = -99999.
    -----
    |                                     |.
    -----
    REPEAT
        Readln(tty); Read(tty, Num);
        Sum = Sum + Num;
        Count = Count + 1;
        UNTIL Num = 99999;
        Average = Sum/Count;
        Writeln(tty, Average);
    END
END

```

	Correct		Incorrect	
	mean	time n	mean	time n
Version Alpha	2.26	37	2.41	10
Version Beta	3.98	3	2.42	41

(the time is in minutes)

**Figure 2:** FIB programs





Cerebral Lateralization, Preferred Cognitive Mode and Reading  
Achievement in American Indian Bilingual Children

Fairlee E. Winfield, Ph.D.  
Northern Arizona University

Although it is generally accepted that the left cerebral hemisphere is dominant for language in most people, the nature of the relations among cerebral dominance for language, academic performance, and preferred cognitive mode remains controversial. As early as 1937, Orton suggested that a strong left hemisphere dominance would be reflected in excellence of verbal functions such as reading.<sup>1</sup> However, cerebral dominance studies conducted with bilinguals seem to indicate that language for them may be more diffusely organized. Albert and Obler proposed a stage hypothesis in which there was participation of the right brain hemisphere in the initial stages of learning a second language with increasing left hemisphere participation as proficiency developed.<sup>2</sup> Unfortunately, further research evidence and a comprehensive review of over one hundred studies of the bilingual brain by Galloway in 1982 have failed to support the stage hypothesis.<sup>3</sup> Nevertheless, it seems evident that bilinguals do show greater right hemisphere participation in language, but how this is related to cognitive style preference and academic performance remains a question.

American Indian bilingual groups have consistently shown a greater right hemisphere involvement in language.<sup>4</sup> Rogers et al. propose that the structure of some Amerindian languages corresponds to an appositional mode of thinking associated with the right brain hemisphere. For example, there is a greater concern with concrete notions rather than with abstractions; there is a projection of subjective experience to outside influences (i.e., I am drowning → water is drowning me); and modified verbs and participles rather than nouns and adjectives assume importance. Navajo handling verbs require a selection process whereby the verb stem differs if the object referred to is long, small, rounded, clustered, bundled up, alive, or inanimate. The Navajo speaker must be fully aware of the perceptual field and concentrate on the physical characteristics of the object rather than its name. In English we say, "Give me the rock, stick, paper, etc." In Navajo a speaker must decide on the object's characteristics before selecting a verb form:<sup>5</sup>

shaa nī'aah \* single, hard, bulky object  
shaa nī'tiih \* single, slender, stiff  
shaa nī'tsoos \* single, flat, flexible  
shaa nī'lé \* single, slender, flexible  
shaa nījāāh \* plural, profusion

shaa nī'jooY \* non-compact matter  
shaa nī'tleeh \* mushy matter  
shaa nī'yeeh \* load, pack, burden  
shaa nī'nīfīY \* plural, slender, flexible  
shaa nī'teeh \* single, animate

Because of the appositional characteristics of Navajo, it is hypothesized that Navajo-English bilingual children will show a greater right hemisphere participation in language than is generally found in monolingual children.

A second notion that bears upon bilingual cognitive strategies is offered by Galloway who sees language as consisting of two factors: linguistic competence (grammar) and communicative competence (nonverbal, interactional, context-dependent abilities).<sup>6</sup> The communicative competence is highly dependent upon neural substrates lateralized in the right brain hemisphere such as perception of facial affect, tone of voice, and linguistic gesture. Galloway proposes that there is a relationship between minority group membership and patterns of cerebral asymmetry. She states that members of bilingual minorities who are unacculturated and socially less powerful in the surrounding Anglo majority world but who are under pressure to assimilate will show greater right hemisphere involvement in language processing. The suggestion is made that the person who must function in two or more languages may have to develop a greater pragmatic flexibility and sensitivity to interactional cues in order to support and maintain communication. No claim is made that the actual left lateralization of the linguistic grammar is altered or that the brain has such plasticity that

the linguistic grammar is suddenly shifted toward the right. Right hemisphere language would be a reflection of a cognitive strategy adopted to increase available communicative cues.

This study examines the relations among sex, school grade, verbal dichotic listening asymmetry, cognitive strategic preference, and reading ability in Navajo-English bilingual children and attempts to compare the findings with those of Caplan and Kinsbourne in their research with monolingual children.<sup>7</sup> Therefore, the following hypotheses are also investigated:

1. Bilingual children who are more highly left hemisphere lateralized for language on the dichotic listening test will show a stronger preference for verbal strategies on cognitive style measures.
2. Bilingual children who prefer the verbal mode on cognitive style measures will demonstrate reading achievement superior to that of the children whose preference is for a nonverbal mode of information processing.
3. Children in the lower grades will display a greater preference for nonverbal strategies than older children.
4. Females will show a greater preference than males for verbal strategies and will display reading achievement superior to that of males.

The subjects were 78 Navajo-English students in an isolated Bureau of Indian Affairs boarding school on the Navajo Indian Reservation. All instruction at the school is in the second language (English); however, all social interaction is in Navajo. The students are from the most traditional area of the reservation where pressure for assimilation is felt, but where there has been relatively little acculturation. The group included 35 boys and 41 girls in grades 3 to 8 (23 in grades 7-8, 27 in grades 5-6, 28 in grades 3-4). The mean age was 10.6. An eight-item handedness measure was administered in both languages to eliminate subjects with a left hand preference.

### Test Instruments

#### Dichotic Listening (Navajo Word Test and English Word Test)

An 80-item Navajo word test and a 74-item English word test was administered to each student. Word pairs were matched for syllable length. Six practice items familiarized the children with the procedure. Headphones were reversed to control for variation in channel intensity. Half the children heard the Navajo test first and half the English test first. The children were told to report both words if possible.

#### Cognitive Mode (Auditory and Visual)

**Word-Shape Sorting:** This test was used in the research of Caplan and Kinsbourne in 1981. The child selects the one item of three that does not fit with the other two. The two equally correct responses indicate a preference for visual-spatial processing (on the basis of shape) or visual-verbal processing (on the basis of words). The test was administered in English only as the students were not literate in Navajo.

**Auditory Style Test:** This test was based upon one used by Gross, Franko, and Lewin (1978).<sup>8</sup> The child selects two verbal items that "go together" from three words heard. The two equally correct responses indicate a semantic preference (synonyms or category relationship) or a nonsemantic preference (rhyme). This test was administered in both Navajo and English since it is entirely verbal.

#### Reading Achievement (California Test of Basic Skills, CTBS)

A reading achievement score was obtained from students' performance on the California Test of Basic Skills that was administered to all of the boarding school students immediately prior to this project. The test is entirely in English and the bilingual students who are from a minority language and culture consistently perform below grade norms. For example:

Grades	Grade Mean	Reading	CTBS Math	Combined
3-4 (n=28)	3.35	2.38	3.02	2.55
5-6 (n=27)	5.44	3.57	4.51	4.04
7-8 (n=23)	7.30	4.48	6.19	4.67

### Results

The data for the total sample were analyzed by correlational methods, ear-differences on dichotic listening were determined by t tests, and sex and group (3-4th grade, 5-6th grade, 7-8th grade) comparisons were made by analysis of variance.

A strong right-ear advantage, indicating left hemisphere lateralization for language, was obtained on the English dichotic listening task for the entire sample ( $t = 3.56$ ,  $p < .001$ ). However, for the Navajo dichotic listening task, there was no significant difference in right and left ear scores indicating a greater right hemisphere participation in Navajo. Comparisons of the Navajo and English ear difference scores indicate that although there is relationship between the two measures ( $r = .329$ ,  $p < .002$ ), students are significantly less left hemisphere lateralized for Navajo than for English ( $t = -2.32$ ,  $p < .02$ ). When the three age groups are considered, there was no significant developmental increase in the size of the right ear advantage for either language (Navajo,  $F = .679$ , N.S.; English,  $F = .058$ , N.S.).

Table 1  
Correlation Matrix, Total (N = 78)

	Age	Navajo Lateralization	English Lateralization	Visual-Verbal Preference	Auditory-Verbal Preference/English	Auditory-Verbal Preference/Navajo	CTBS Reading
Age		-.0946	-.1025	.1924*	.2901**	.3732**	.6616**
Navajo Lateralization			.3297*	-.2340*	-.0380	.1113	-.0501
English Lateralization				.0308	-.0203	.0407	-.0528
Visual-Verbal Preference					.1761	.0803	.4430**
Auditory-Verbal Preference/English						.6396**	.2508*
Auditory-Verbal Preference/Navajo							.3585**
CTBS Reading							

\* $p < .05$ .

\*\* $p < .01$ .

Correlational data for the total sample is presented in Table 1. There is a highly significant correlation between strength of verbal performance on all visual and auditory cognitive style measures and performance on the CTBS reading measure. However, there is no indication of a relationship between the visual and auditory cognitive style measures indicating that they may be measuring separate aspects of cognitive style preference. Dichotic listening measures are unrelated to reading performance, but the Navajo dichotic listening measure is negatively correlated with performance on the visual-verbal preference measure.

No sex differences for verbal preference were found; however, a significant developmental increase in verbal preference on the auditory cognitive style measures was obtained for the sample as a whole (Navajo-Auditory,  $F = 4.506$ ,  $p < .01$ ; English-Auditory,  $F = 7.131$ ,  $p < .01$ ). No developmental increase occurred for the visual-verbal preference measure. Significant developmental increases were evident for the CTBS reading scores as was anticipated.



Given that verbal strategic preference correlates significantly both with age and reading ability, and the age-related improvement in reading, the association between cognitive style and reading could be a consequence of a common association with age. Therefore, as in Caplan and Kinsbourne's analysis, partial correlations were calculated to eliminate reading age (grade equivalent) and chronological age.<sup>9</sup> Results for the English language cognitive style measures are consistent with previous findings that greater preference for verbal strategies accompanies better reading, regardless of age. However, this does not hold for the Navajo cognitive style measure where greater preference for verbal strategies does not accompany better reading but is based upon age and reading proficiency.

## Discussion

Caplan and Kinsbourne suggest that the nature of the problem-solving strategy adopted may affect the size of the observed lateral bias. If a strategy is selected which has a predominantly lateralized neural substrate, then one hemisphere will be preferentially activated. However, some tasks may be adequately accomplished by more than one strategy and a strategic choice may be relatively stable across situations regardless of cognitive demands of the task.<sup>10</sup> Therefore, it is possible to speculate that Navajo-English bilinguals, although anticipating verbal input in Navajo, may preferentially engage the right hemisphere and adopt a holistic, pragmatic, interactional strategy biasing attention to the left side of space. The degree of dichotic listening asymmetry becomes an index of differential hemispheric use without structural implications for the degree of lateralization of function. Initial preference for a particular mode may become habitualized, and there is some evidence that stylistic preferences may be learned.<sup>11</sup> If such is the case, the greater left lateralization for English may be a product of the classroom situation that is heavily biased toward a verbal mode with sequential, analytic processing.

The present finding of right hemisphere representation of language may be a reflection of the habitualized use of an interactional, nonverbal strategy to obtain additional cues from language input. Since the dichotic listening task is unfamiliar, the students may shift to a right hemisphere strategy to try to pick up additional cues. They may indeed not recognize the situation as a language task in Navajo while they assume all situations involving English are language tasks because of habitualized classroom experience.

Since the preference for verbal strategies was unrelated to the size of the left hemisphere lateralization, and there is no positive correlation between results on dichotic listening and verbal cognitive style preference, the current findings would support the notion that dichotic asymmetry is an index of differential hemisphere use without structural implications of function. Additionally, since asymmetry did not increase significantly with increasing age, results support earlier studies that argue for the existence in infancy of hemispheric specialization and stability of lateralization across ages.<sup>12</sup>

No sex differences on lateralization, cognitive style, or reading achievement were evident for the bilingual students although previous studies have generally reported greater verbal preference and advanced reading scores for girls and frequently stronger right ear advantages for boys. It is speculated that the English language barrier, Navajo cultural expectations, and the factor of minority group membership tend to override sex differences.

The pivotal variable, preferred cognitive strategy, was found to relate to reading performance. Preference for a verbal approach on both the visual and auditory cognitive mode tests is associated with better reading and the association is maintained when effects of ability and age are eliminated. Caplan and Kinsbourne suggest that performance is improved when task demands and cognitive bias are concordant. Therefore, a child who typically chooses verbal strategies will more easily grasp the nature of the "reading code."<sup>13</sup>



STRONG COGNITIVISM  
CONTRARY PCNUSM

M  
Graham, Stephen