OUTSTANDING QUESTIONS IN COGNITIVE SCIENCE

A SYMPOSIUM HONORING TEN YEARS OF THE DAVID E. RUMELHART PRIZE IN COGNITIVE SCIENCE

August 14, 2010 Portland Oregon, USA

David E. Rumelhart

ABOUT THE RUMELHART PRIZE AND THE 10 YEAR SYMPOSIUM

At the August 2000 meeting of the Cognitive Science Society, Drs. James L. McClelland and Robert J. Glushko presented the initial plan to honor the intellectual contributions of David E. Rumelhart to Cognitive Science. Rumelhart had retired from Stanford University in 1998, suffering from Pick's disease, a degenerative neurological illness. The plan involved honoring Rumelhart through an annual prize of $100,000, funded by the Robert J. Glushko and Pamela Samuelson Foundation. McClelland was a close collaborator of Rumelhart, and, together, they had written numerous articles and books on parallel distributed processing. Glushko, who had been Rumelhart's Ph.D student in the late 1970s and a Silicon Valley entrepreneur in the 1990s, is currently an adjunct professor at the University of California – Berkeley.

The David E. Rumelhart prize was conceived to honor outstanding research in formal approaches to human cognition. Rumelhart's own seminal contributions to cognitive science included both connectionist and symbolic models, employing both computational and mathematical tools. These contributions progressed from his early work on analogies and story grammars to the development of backpropagation and the use of parallel distributed processing to model various cognitive abilities. Critically, Rumelhart believed that future progress in cognitive science would depend upon researchers being able to develop rigorous, formal theories of mental structures and processes.

Fitting with Rumelhart's own interdisciplinary and multipronged attack on the problems of cognitive science, the first ten recipients of the Prize have come from a number of fields (computer science, linguistics, and psychology), tackled a number of core issues (learning, development, vision, language, induction, generalization, and memory, to name a few), and used a wide variety of theory-building tools (including process models, optimality analyses, statistical modeling, neural networks, and knowledge representation).

In this symposium, the concluding session of the 2010 Annual Meeting of the Cognitive Science Society, the first ten recipients of the Rumelhart Prize pose some important unanswered and perhaps even unasked questions for cognitive science. In so doing, the previous prize winners are formulating challenges that await the future prize winners.

The symposium begins with a 1987 video in which David Rumelhart explains the nature of the challenges that framed his own research. The transcript is included here.

- William Bechtel, Symposium Moderator
- Robert Goldstone, Chairman, Rumelhart Prize Selection Committee
- Robert J. Glushko, Glushko-Samuelson Foundation

TABLE OF CONTENTS

David E. Rumelhart

[Note: The following is an excerpt from the recorded question period following the presentation of a lecture by David E. Rumelhart on Brainstyle Computation and Learning, a lecture in The Stanford Computer Science Video Journal: Artificial Intelligence Research Lectures, San Mateo, CA: Morgan Kaufmann. Copyright © 1987 by the Leland Stanford Junior University. Transcribed and lightly edited by James L. McClelland, July 25, 2010.]

Question:

What are the main problems that need to be solved in the PDP approach?

Rumelhart's response:

I can tell you the things that I intend to be working on, which were chosen in part because they seemed like important problems. One issue has to do with what you might call higher-level reasoning. I'm interested in trying to give an account of how people can do logic, how they can do mathematics, and so on – things that seem rather foreign, at first blush, to these kinds of systems. I believe that there are a number of features of these systems that will allow them to behave more like people in the face of these sorts of things.

One area I think is important is the case of analogical reasoning, something I think these systems are going to turn out to be well suited for, indeed I've done some work on this. It's been more traditional let's say in AI to start with logic as your prime thing and so doing logic was never a problem because that was the basis of it; whereas in this case, we've more or less turned this on its head. We've said no, logic isn't the center of this kind of a computational system, so then we can say how then do people do logic? Well my answer to that is to say: not very well; but nevertheless they can – they can learn to do it, and the question is, how is that accomplished by these kinds of systems?

Another question has to do with how we deal with highly structured entities like sentences and pieces of discourse, which are highly structured pieces of data that, in principle, are unbounded – there's no finite length. The issue is that you don't want any a priori restrictions of this kind. How do you deal with those kinds of things? Those are two central questions.

Geoffrey Hinton (2001)

Where do features come from?

In the Hitchhiker's Guide to the Galaxy, a fearsome intergalactic battle fleet is accidentally eaten by a small dog due to a terrible miscalculation of scale. I think that a similar fate awaits most of the models proposed by Cognitive Scientists.

Much of Cognitive Science is attempting to explain extremely complex, emergent phenomena using almost no mathematics and with very little understanding of the computational abilities of the neural networks that give rise to these phenomena. I find a comparison with solid state physics quite illuminating. Systems of interacting spins exhibit interesting emergent phenomena called "phase transitions'' when all of the interaction strengths are gradually scaled up by exactly the same amount. In these systems, the emergent phenomena are the result of changing one scalar without any feedback. It has taken a lot of smart physicists quite a long time to develop a good understanding of these systems and they are still not finished.

In the brain, the interactions between neurons change in a complicated way that depends on their current values and also depends on the inputs and outputs of the system. Is there any reason other than pure hubris for thinking that the phenomena that emerge from such complex changes can be understood without developing a huge amount of mathematical understanding?

The questions posed by Cognitive Scientists are of the utmost importance, but a real answer to most of them will only be obtained by understanding how the system actually works and this requires answers to much more basic questions. Without that basic understanding, the answers that Cognitive Scientists agree on will be no more correct than most people's opinion of why it’s so much easier to hear a conversation when you are downwind of it.

So what is an example of one of these basic questions? Research on pattern recognition strongly suggests that good recognition requires good features. There are several ways in which such features could be discovered. Evolution might stumble across them using a search technique that is really bad at exploring high-dimensional spaces. They could be developed by back-propagating recognition errors, had we but labels enough and time. They could be created by a silly vector method that stores entire sensory input vectors and treats them as features that are compared with new inputs using a fixed comparison metric. Or they could be learned by fitting a multilayer generative model to the sensory input.

Richard M. Shiffrin 2002

What is context and its role in memory and learning?

My first reply to Bob Glushko, somewhat-tongue-in-cheek but in good part real, was my wish to understand the workings of mind and brain. This question being slightly too broad to spark serious discussion, I instead ask about memory, which underlies every cognitive activity and defines who and what we are. Memory has been the subject of intense speculation and study since the first days of philosophy, psychology, and neuroscience, but one cannot help harboring the suspicion that we are all blind men exploring different parts of the elephant. Understanding memory requires knowledge of chemical, biochemical, and physical mechanisms in the smallest and most primitive living organisms, of processes and processing networks in non-human organisms of all sorts, and of chemical, neural, network, behavioral, emotional and cognitive complexities of human beings throughout development.

Such questions being far too broad, let us restrict inquiry further: E.g. How does general knowledge form from the events experienced, and how does the current form of personal knowledge affect the coding of events? What role does the hippocampus (and other brain structures) play in the answer to these questions? Given changing knowledge and memory as inference and reconstruction, is there any way to assess the permanence of event memory? To what degree are memories stored separately or in composite form, either neurally or functionally, and does the answer differ for event memories and knowledge? What role is played by organized phasic firing of large networks of neurons, and how does this interact with both synaptic changes and behavioral storage and retrieval? Where does consciousness arise and what is its role in memory? How much of memory storage and retrieval occurs automatically and how much is guided by attention and similar processes (more sensibly we should ask how these interact)? What is the status of different memory systems, often characterized roughly in terms of retention duration; e.g. sensory memories, working memories, short-term memory, active memory, long-term memory, and knowledge?

All this remaining far too much, I posed the question: “What is context and its role in memory and learning?” Context often refers to any sort of information not controlled or varied by an experimenter. As such, there exists the danger that it is used as a theoretical sponge to absorb all failures of one’s favorite theory. Somewhat recently, progress has been made by decomposing context into its component parts and subjecting those components to empirical test. Nonetheless, although context remains a critical construct of most theories, there remain numerous questions: What kinds of contextual information are most important? How much of contextually based storage and retrieval is under control and how much automatic? What happens to context as knowledge is formed? How does context change as time and the environment change? What role does context play in language, and what types of context are critical for language? Does context provide the underpinning for embodied cognition? To what degree is old context reconstructed when retrieving? What are the neural and hippocampal underpinnings of context in storage and retrieval?

Aravind Joshi 2003

What is the nature of Immediate Discourse (ID) and the transition from sentence to ID?

It is not hard to pose questions about language and cognition, whose answers we all would like to know. These are questions about the relationship of language and other cognitive systems and they are such that it is easy to pose the questions but hard to speculate when we might see some real breakthroughs, perhaps over a horizon of 25 years or longer.

I would like to pose a particular question, which is related to a very important aspect of language, yet constrained enough such that we can hope to get some very insightful answers in a period of 10 years or so.

For the ease of discussion, I will characterize this issue as the nature of Immediate Discourse (ID). Besides some prescriptive and normative guidelines, we still do not really understand (a) what takes us beyond a sentence into discourse, (b) what aspects of sentence structure carry over into discourse, (c) how discourse aspects affect sentence structure, etc. Most of the theoretical and experimental work concerning language is very largely concerned with what happens in the domain of the sentence, i.e., it is primarily sentence based. Although some appeal to context is often made, the work tends to be essentially bound by the notion of a sentence.

One key aspect of sentence structure is predicate-argument structure as represented by the syntax and the compositional aspects of semantics at the sentence level. By Immediate Discourse (ID) I mean the aspects of discourse structure that can be characterized by lifting these sentence bound notions to the discourse level. The predicate argument structure exhibited by so-called discourse connectives, expressed either explicitly (through conjunctions, adverbials or phrases that are neither) or implicitly, provides a way of defining ID. By immediate I do not mean some bounded piece of discourse beyond a sentence but rather that aspect of discourse which can be accessed via connectives and thus associated with compositional semantics. It is unbounded in a sense. I use the term immediate to distinguish it from other aspects of discourse such as intentional structure, overall plan, etc.

I believe that a systematic attack on the problem of transition from sentence to ID will provide insights into the nature of ID with direct relevance to cognitive science. For example, (a) we might get elegant answers as to why sometimes at the ID level some of the work done by syntax at the sentence level needs to be undone, (b) to what extent the senses of the connectives are stable across languages, (c) what are the different ways these senses get expressed in a particular language and also cross-linguistically, (d) is the inventory of these expressions closed or partially open and if so to what extent, etc. All these issues, I believe have direct relevance to our efforts in understanding the various aspects of the language faculty.

John R. Anderson 2004

What will Cognitive Science be like after the Singularity?

Psychology has always striven to take the complexities of human experience, abstract away, and infer relatively simple principles that capture significant portions of variance. This has meant that we have thrown away "as noise" peculiarities of different individuals and their moment-to-moment experiences. There were two principal motivations for this:

1. It is inherent in the concept of "understanding" and "explanation" that the complex is reduced to the simple.
2. We did not have the machinery to make testable predictions that dealt with all that detail of human experience.

While the first argument remains valid, the second is being challenged by the continued advancement of computational power. We increasingly have the prospect of being able to predict what we do not understand. While the applications of this potential are many, I will focus on applications to education where the goal of adapting instruction to the individual student has largely remained an empty platitude. I will illustrate with some incipient work that uses brain imaging to read the minds of students as they solve mathematical problems. While we have had some success at doing this, we do not have a full grasp on how the program is doing it.

Paul Smolensky 2005

Can we stop denying and start solving the host of cognitive problems in which continuous and discrete aspects of representations strongly interact, now that we know how to construct Parallel Distributed Symbol Systems?

What are the cognitive problems?

LANGUAGE: EXAMPLES

- Phonetics and phonology respectively use continuous and discrete characterizations of knowledge, but it is widely recognized that there is a great deal of overlap in the substance of this knowledge. Can we build a unified theory?
- In phonological production, continuous activation-spreading constructs outputs that are discrete, to a good approximation, but also gradient in subtle ways. Can this be accounted for within a single integrated architecture?
- In many aspects of linguistic performance, continuous variables such as frequency and similarity interact strongly with discrete grammatical structure. Can we derive such interaction from the cognitive microstructure of grammar?

LEARNING: EXAMPLES

- Does learning, in development and adulthood, construct general symbolic knowledge systems (such as the hypothesis spaces of many Bayesian models) through subsymbolic statistical learning?

- What are the implications for Bayesian learning models of an underlying distributed microstructure?
- Can a single, general-purpose subsymbolic substrate for symbolic processing suffice for acquiring symbolic knowledge in all cognitive domains? Is the structure of such a substrate learnable? Is it shared with non-human animals?

What are the methods of solution?

MODELING: BEYOND "CONNECTIONISM VS. SYMBOLISM"

- Parallel Distributed Symbol Systems:
  – are PDP networks that can be formally described as computing functions over symbol structures;
  – model components of higher cognition at a granularity where processing is primarily parallel;
  – are linked through serial and parallel processing into larger-scale architectures.
- Symbols are realized as distributed activation patterns, which live in a continuous representational space where they undergo fundamentally continuous processing; they are related by continuous relations such as similarity, relative activation, and frequency of instantiation.
- Crucially, however, the representational space is also combinatorial.
- These models provide strong hypotheses about what to look for: Can we find grammars in brains?

AXIOMATICS: BEYOND MODELING

- Mathematical analysis of the PDP microstructure yields principles for new types of symbolic theory. Optimization properties, for example, percolate up from the subsymbolic micro-structure to the symbolic macrostructure, yielding optimality-based grammars. What other properties percolate up?

- At the core of theories of cognitive macrostructure are many abstractions: representation, hierarchy, mental activation, functional architecture. Can such abstract notions be made formally precise through axiomatization?
- Can Axiomatic Theory In Cognitive Science—AxiomaTICS—enable formal derivation of experimental predictions from detailed hypotheses about the structure of (e.g., linguistic) mental representations? Intuition does not provide a reliable means of making such predictions.
- The AxiomaTICS level abstracts away from many microstructural details of the PDP modeling level, but infuses symbolic theory with continuous structure arising from this microstructure (e.g., similarity of different symbolic roles). Can AxiomaTICS provide new explanations of competence- and performance-level empirical generalizations?

How general is the method?

CULTURE: BEYOND NARROW COGNITION

- Are optimization-based grammars widespread in higher cognition, beyond linguistic knowledge?
- Can they be used to explain universals and cross-cultural typologies of, e.g., moral systems, kinship systems, conceptual systems?

Roger Shepard 2006

Can cognitive science provide rational grounds for universally valid moral principles?

Or do moral judgments stem solely from evolutionarily and culturally shaped interests in one's own self, kin, tribe, or species? I am seeking support for the former possibility.

Humans may be the most cognitively advanced and rational of terrestrial animals. But cognitive biases, erroneous premises, and logic-tight compartmentalizations still enable primitive and destructive motives of greed, bigotry, racism, and retaliatory rage.

Some progress is nevertheless discernable. Slavery and denial of women's right to vote (both tolerated even by our democracy's "founding fathers") are now recognized as inconsistent with rational moral principles. And human cognition has proved capable of developing rational (logical, mathematical, and scientific) tools for discovering and transcending its own imperfections.

I proposed that rationally-based moral principles are universal in that they would be embraced by any cognitive beings that have sufficiently achieved what I termed "the step to rationality"—regardless of where such beings evolve in the universe. As indicated in my 2008 Cognitive Science article "The Step to Rationality," this proposal grew from cases I had developed for "universal mental laws" such as those of generalization and mental transformation.

I argued that our deeply internalized capability of mental transformation underlies an intuitive grasp of the symmetry of invariance under permutation of individuals—the abstract principle behind empathy, Kantian universalizability, Rawlsian justice, and the "Golden Rule." Further, I offered a cognitive account of how the free will required for the attribution of moral responsibility is compatible with determinism. But a cognitive theory of morality must also distinguish two classes of "individuals": one for those deemed to have moral rights, the other for those held to be morally responsible.

The class of responsible individuals I (like Kant) take to be the class of rational beings. Among terrestrial animals, this is presumably the class of (sufficiently rational) human adults. But, the boundary between these and children, and the brain damaged, demented, or psychotic is not sharp. So, the definition of the equivalence class of rational individuals remains fuzzy. (In courts of law, it may reduce to an intuitive judgment of whether a defendant can distinguish right from wrong.)

The class of individuals with the right to be treated morally by the morally responsible beings I (like Bentham) take to be the vastly larger class of sentient beings. Among terrestrial animals, this includes all individuals capable of experiencing pleasure and pain, well-being and suffering. But there is no definable boundary here, only a gradual fall-off of degrees of sentience (as might be inferred from behavioral, neurophysiological, and genetic information about various species).

Any attempt to formalize rational moral principles concerning rights and responsibilities, alike, therefore requires a significant extension of my initially proposed invariance permutation of individuals in a well-defined equivalence class. We need to accommodate suitably objective, morally justifiable, and quantitative degrees of class membership and, hence, of (possibly asymmetric) exchangeabilities of individuals.

Jeffrey L. Elman 2007

How does development work? And why is this such an important question for cognitive science?

The study of development is hardly a novel topic. But the potential of the study of development to shed light on a number of core questions in cognitive science is not widely appreciated (or not as appreciated as I will argue it should be). Below I point to four questions, which may be best answered by studying development. I am certainly not alone in noting these implications, but I think the field has yet to appreciate their significance. My role here will thus be to play booster for the study of development.

The origins of complexity

The typical approach to understanding complex cognitive systems is to study them in their mature state. Reverse engineering and decompiling may work in some cases, but I suspect that for truly complex mechanisms, the only feasible approach is to focus on the process by which a mechanism grows and becomes complex. The study of development is thus not the study of immature systems but the processes of change that transform (in the case of humans) single cells into beings with trillions of cells whose organization and behavior is extraordinarily complex.

Development as a constraint on theory

A critical assumption of the behaviorist enterprise was that one cannot understand internal psychological processes because, being unobservable, theories would be unconstrained. The study of cognition became possible when it was recognized that insights from multiple disciplines (neuroscience, automata and information theory, anthropology, computer science, etc.) could provide sufficient constraints to make theorizing feasible. If one also assumes that development is a continuous process, the study of development provides a powerful additional constraint ontheories of adult systems.

The interaction of mechanism of context

Although the 'Nature vs. Nurture' question is freighted with a history of poorly posed alternatives, there is no question that the development of biological systems reflects a multitude of two-way interactions between the system's intrinsic properties and the environment in which it matures. The organism-context interaction in development may be one of the most stunning 'context effects' in nature.

Understanding learning

Learning has been a central topic in cognitive science, particularly in recent decades, and we have made tremendous progress in this area. Nonetheless, a feature of most models is that the learning mechanisms are viewed as largely stationary (in the technical sense that their statistical properties and processes do not change over time). But the stunning fact about human learners is that they learn the most at precisely the point in their life history when they are themselves undergoing tremendous change. What can be learned at any point in time both reflects the state of the system at that time, and also sets the stage for what can be learned in the future.
We understand almost nothing about the ways in which development and learning interact.

Shimon Ullman 2008

What innate structures will a 'digital baby' need to acquire meaningful concepts from its perceptual experience?

A basic question for cognitive science I would like to pose is a theory of 'computational Nativism' – a computational theory of cognitively and biologically plausible innate structures, and how they guide the cognitive system along specific paths through its acquisition of knowledge, to continuously acquire meaningful concepts and useful representations.

In computational modeling of learning from visual experience, the focus to date has been mostly empiricist in nature. Probabilistic models have been developed as powerful tools for extracting the unobserved causes of sensory signals. Such models can efficiently discover significant statistical regularities in the observed signals, which may be subtle and higher-order, and construct worldmodels based on the estimated probability of the observed data. However, statistical empiricist learning is insufficient for learning to understand the world, including cognitive semantic categories and concepts, which depend not only on statistical regularities in the sensory input, but also on their significance and meaning to the observer. As has been shown by a rich body of developmental studies, the human cognitive system is equipped through evolution with basic innate structures that facilitate the acquisition of meaningful concepts and categories. These are used to obtain a 'true understanding' of the world, which goes beyond correlations and statistical regularities.

I envision a 'digital baby' model which, through perception and interaction with the world, gradually develops on its own representations of complex concepts that allow it to understand the world around it, in terms of objects, categories, events, agents, actions, goals, social interactions and the like. The model should incorporate a set of innate domain-specific concepts, priors, policies and capacities, which will guide the developing system through its subsequent acquisition of knowledge in these domains. The innate knowledge is likely to be primarily not in terms of developed concepts, but in terms of 'proto-concepts', which serve as anchor points and initial directions for the subsequent development of mature concepts.

Some of the challenge of 'computational nativism' will be to study by analysis, simulations and testing, plausible innate structures, which, on the one hand, are consistent with empirical data, and, on the other hand, can actually 'do the job' in the sense that incorporating them into the learning system can demonstrably lead to a semantic level of representing the world which goes beyond the reach of current approaches. In addition to their cognitive role, challenging questions regarding the innate cognitive structures are their evolutionary development, their genetic encoding and their embodiment in the brain.

Susan Carey 2009

What process(es) underlie the creation of new concepts?

The problem: Only humans can think thoughts formulated in terms of most of the concepts lexicalized in natural languages. Other animals cannot ponder the causes of cancer, or whether there are more real than rational numbers. Nor can human infants. How can the disciplines of cognitive science understand the acquisition of the human conceptual repertoire?

Unpacking the question: The British empiricists (who have many heirs among cognitive scientists today) held that our concepts bottom out in a set of representations that are primitive in three different senses: they are developmental primitives (they are the output of innate input analyzers), they are interpretational primitives (they have innate conceptual roles and need not be decomposed into other representation to underlie thought), and they are definitional primitives (all concepts can be formulated in terms of them, either in the form of data structures that provide necessary and sufficient conditions for category membership or as data structures that determine reference probabilistically, as in prototype or exemplar theories). The empiricists believed that the same set of primitives (in their theories, sensori or sensori-motor primitives) played all three roles, again a position with adherents to this day. This position is easily dismissed. Today we can ask, what is the atomic number of gold? Indeed, in today's chemistry, "element with atomic number 79" can be taken as a definition of gold. But where did the concepts "element" or "atomic number" or even "79" come from? These are not innate primitives, nor can they be expressed in terms of sensori or sensori-motor primitives (as the logical positivists learned at great cost to their program of research).

My concern here is with developmental primitives, as well as what we might call "learning primitives." Part of the project for cognitive science is discovering what the innate representational resources are—what representations are input to learning mechanisms at the very outset of development. But the issue I wish to focus on arises in every episode of learning. All learning mechanisms (from logical hypothesis testing mechanisms, to Bayesian learning algorithms, to various connectionist or associative mechanisms) must posit a set of representations available at the outset, that are the input to the process. The question is how these representations, the concepts that articulate the hypothesis space in Bayesian algorithms, the features that are the input and output representations in back propogation algorithms, arise.

Another way of stating the problem is: if there are discontinuities in the course of conceptual development, how are these to be described (i.e., what kinds of discontinuities are there; in what sense of "discontinuity") and what learning mechanisms can achieve them. There is good evidence for at least two partially distinct types of conceptual discontinuity in the course of the historical development of conceptual representations, each of which must be recapitulated as the child masters these historically created resources: increases in representational power (as in mathematics with the development of integer representations, rational and real numbers, and the calculus), and incommensurabilities, in which new conceptual systems emerge that cannot be expressed in terms of the concepts available at the outset of the conceptual change. Both of these involve the creation of new conceptual primitives.

Another version of the problem: What makes us so smart? The study of perceptual and sensory representations reveals great continuity over evolution; our perceptual systems resemble those of other primates in the computational machinery that underlie them, and in their neural substrate. Recent work extends this generalization to more conceptual representations—innate representations of number, agents, and the physical world are shared with other primates (and in many cases, with other mammals or vertebrates more widely.) These are evolutionarily very ancient representational systems. But we are the only animal who has discovered the atomic number of gold. How do humans culturally construct new representational systems, formulated in terms of concepts not previously available, and how do children acquire these new systems of representation, given that they are articulated in terms of primitives they do not yet represent?

James McClelland 2010

How do we know when we don't understand, and how do we find useful steps toward understanding?

Many times in my life, I have had the experience of seeking to understand something I did not already understand. This sort of experience frequently happens in science, and also in everyday life, when one seeks to understand some observed event or phenomenon.

One can, of course, have the illusion of understanding. People used to think they understood the process that gave rise to the properties of different plant and animal species, or to particular states of the world such as successive years of famine: these things arose due to the actions of the gods. But there came a time when some thinkers and scientists began to find such answers wanting, and struggled to find alternative explanations. There are many other examples fully within a scientific framework, when one theory or approach no longer seems adequate and one begins to seek a new understanding.

Many of the supreme achievements of the human mind begin with a feeling of dissatisfaction with conventional understanding. This was certainly the case with Darwin, who realized that neither the theological nor the quasi-scientific ideas about the basis of evolution that were in play at the time were adequate. This led him to spend many years trying to work out a satisfactory alternative to conventional wisdom about the origin of species. Darwin's first insight was to realize there was a problem to be solved. But this realization left him a very long way from the solution to his problem. What is fascinating in the story of

Darwin's development of the idea of natural selection is that he recognized the relevance of certain ideas, which, by themselves, were not fully adequate but which, when combined with others, led to his theory of natural selection. One such idea was the Malthusian notion that populations always expand until they reach the limits of available resources, at which point not all members can survive. In one biography of Darwin (D. Quammen, The Reluctant Mr. Darwin), it is stressed that he recognized the relevance of this idea to the question of evolution, without yet recognizing how limited resources would lead to a process of selection.

A future challenge for our field will be to try to understand how Darwin and other scientists successfully complete the arduous journey from dissatisfaction to understanding, first recognizing the important gap, then finding useful ideas on which to build, then finally discovering exactly how to build on them to reach a new understanding. Such work will be important, not only for a cognitive science of science, but also for a cognitive science of naturalistic everyday problem solving, where events in nature and social interactions occur that force us to rethink our current understanding. I expect connectionist ideas will be relevant to addressing these issues, but I also expect a solution to require extension of the connectionist approach and/or draw on other frameworks and approaches.


GLUSHKO-SAMUELSON FOUNDATION