

# Organization and Emergence of Semantic Knowledge: A Parallel-Distributed Processing Approach

**James L. McClelland (jlm@cnbc.cmu.edu)**

Department of Psychology and Center for the Neural Basis of Cognition  
Carnegie Mellon University, Pittsburgh, PA 15213

How do you know that Socrates is mortal? More generally, how do you know what properties to attribute to an object? How is the relevant knowledge acquired? How is the knowledge organized in the brain, and how is it affected by brain damage? My colleagues and I have been seeking to answer such questions by developing computational models of semantic cognition and its development.

## Parallel-Distributed Processing

Our overall framework relies on an approach to semantic cognition first suggested by Hinton (1981). Hinton's proposal was that our knowledge of the properties of objects as expressed in propositions about them, such as 'A canary can fly', is not stored directly in propositional form but in the strengths of connections between simple processing units that allow propositions to be formed mentally by pattern completion – e.g. filling in a pattern of activation representing 'fly' when probed with patterns representing 'canary' and 'can'. After the introduction of the back-propagation algorithm, Hinton (1989) and Rumelhart (1990; Rumelhart and Todd, 1993) showed how this learning algorithm could discover useful internal representations that would support generalization. In work building on their efforts, my colleagues and I have proposed an overall architecture for the representation, learning, and processing of semantic information. This architecture provides the context in which we have gone on to address questions like those enumerated above.

## Complementary Learning Systems

A key element of the approach is that semantic cognition takes place within a distributed network of contributing brain areas that work together to allow us to learn, represent, and process semantic and other types of information (McClelland, McNaughton, and O'Reilly, 1995). One part of this network, the neocortical learning system, allows for the developmental elaboration of organized semantic representations through a gradual learning process. The other, located in the medial temporal lobes, provides a mechanism for learning new information rapidly, while avoiding catastrophic interference that would otherwise occur if the new information were quickly incorporated into the neocortical learning system. For present purposes I focus here on recent work (Rogers & McClelland, 2004) addressing processes we think of as taking place primarily within the neocortex.

## Differentiation and Disintegration of Conceptual Knowledge

In our approach, cognition begins not so much with booming buzzing confusion but with a bland conceptual uniformity. At first, all things are represented with highly similar, undifferentiated patterns of activation. As the network experiences the various properties different things manifest in different contexts, it gradually comes to differentiate them. This differentiation process is sensitive to patterns of coherent co-variation of properties of objects. That is, it picks up on the fact that there are many objects in the world that have wings and feathers and can fly and many others that have four legs, a tail, a wet nose, and can bark. First general and subsequently more specific differentiations occur. The representations of objects reflect their underlying conceptual similarity even as they become more and more differentiated. After learning, progressive damage to the network results in gradual disintegration of the conceptual knowledge, largely reversing the patterns seen in development. These differentiation and disintegration processes closely follow patterns seen in children's cognitive development and in the progressive loss of conceptual knowledge in patients with semantic dementia (Rogers et al, 2005). These effects coexist, in both the model and in human performance, with early emergence of the 'basic level', and with frequency, typicality, and expertise effects.

## Capturing Phenomena Others have Attributed to Innately Constrained Naïve Domain Theories

Beyond capturing developmental and neuropsychological progressions, the model can address several findings others have attributed to naïve domain theories constrained by innate knowledge. These include:

1. Early signs of sensitivity to domain- and property-specific patterns of generalization of attributions from one object to another (Macario, 1991; Gelman and Markman, 1986)
2. Illusory attribution of properties to objects, e.g. attributing legs to animals that do not have them (Williams and Gelman, 1998).
3. Conceptual reorganization and coalescence of categories with the accumulation of experience across varied contexts (Carey, 1985).

The model exhibits these phenomena as a result of gradual learning sensitive to the structure present in its inputs,

unaided by innate domain-specific constraints. Thus the model reopens the question of whether we need to invoke such constraints. It also raises questions about how useful it is to characterize conceptual knowledge as theory-like.

### **Architectural Constraints on the Neocortical Learning System**

The successes of the model are heavily dependent on properties of its architecture: crucially, its reliance on distributed representations constrained to reflect all aspects of the properties that objects exhibit across different contexts. These observations provide the basis for a new way of construing the functions of the anterior temporal neocortex, the region implicated most strongly in the disintegration of conceptual knowledge in semantic dementia. Similarly to Antonio Damasio (1989), we see this region as a 'convergence zone' where from different modalities and different contexts is brought together about the same object (McClelland and Rogers, 2003). For us it is brought together not only to bind the different elements of the conceptual representation, but also to allow the learning process to shape these convergent representations in a way that captures coherent co-variation of properties objects may have across modalities and contexts.

### **Using and Inferring Causal Properties**

Gopnik et al (2004) review an interesting series of experiments on children's ability to attribute causal powers to objects. These authors argue that their findings implicate an innately pre-specified causal inference mechanism that enables children to make such attributions. However, we have implemented a model within our approach that uses the same mechanisms that address the phenomena above to address the bulk of the Gopnik et al. findings. The results suggest that within the domain of causal inference, as well as in other aspects of semantic and conceptual cognition, domain-general mechanisms based on the principles of parallel distributed processing may be sufficient. Surely there are innate biases that guide our semantic cognition, but we suggest that they are domain general principles embodied in PDP networks organized to promote cross-domain convergence of constraints on conceptual representations.

### **Acknowledgments**

Many colleagues have contributed to the ideas presented here, most notably Tim Rogers. The work was supported by a National Institute of Mental Health Interdisciplinary Behavioral Science Center Grant (MH 64445).

### **References**

- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.
- Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation*, 1, 123-132.
- Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, 23, 183-209.
- Gelman, R., & Williams, E. M. (1998). Enabling constraints for cognitive development and learning: A domain-specific epigenetic theory. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology, Volume II: Cognition, perception and development* (Vol. 2, 5 ed., p. 575-630). New York: John Wiley and Sons.
- Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Schulz, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review* 111, 3-32.
- Hinton, G. E. (1981). Implementing semantic networks in parallel hardware. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (p. 161-187). Hillsdale, NJ: Erlbaum.
- Hinton, G. E. (1989). Learning distributed representations of concepts. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology* (p. 46-61). Oxford, UK: Clarendon Press.
- Macario, J. F. (1991). Young children's use of color in classification: Foods and canonically colored objects. *Cognitive Development*, 6, 17-46.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102, 419-457.
- McClelland, J. L. & Rogers, T. T. (2003). The Parallel Distributed Processing Approach to Semantic Cognition. *Nature Reviews Neuroscience*, 4, 310-322.
- Rumelhart, D. E. (1990). Brain style computation: Learning and generalization. In S. F. Zornetzer, J. L. Davis, & C. Lau (Eds.), *An introduction to neural and electronic networks* (p. 405-420). San Diego, CA: Academic Press.
- Rumelhart, D. E., & Todd, P. M. (1993). Learning and connectionist representations. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (p. 3-30). Cambridge, MA: MIT Press.
- Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). The structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, 111, 205-235.
- Rogers, T. T. & McClelland, J. L. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach*. Cambridge, MA: MIT Press.