select subsets for consideration, and these would have to be specified by the theory.

Second, in the model, explanation is instantiated as links between the various units in the network. However, because the psychological processes that underlie explanation are not specified, one could just as well have substituted "connected to" or "associated with" for the term "explains" throughout the article and the model would not need to be changed. An adequate psychological theory must specify the processes involved in "explaining." In certain sciences, researchers spend a significant amount of time determining whether a proposition really explains something, weighing up strong and weak senses of a proposition, and attempting to separate ancillary, ad hoc proposals from the basic tenets of a theory. Thagard hints at this problem with the concept of explanation when he says that a causal sense of "explain" was used in the analyses, but this is inconsistent with his initial statement that the theory should be a theory of any form of explanation. From a modelling perspective, the determination of what it means to explain and what constitutes a theoretical proposition have all the hallmarks of heuristic, evaluative processes. Although such processes could be modelled in a connectionist fashion, a traditional symbolic treatment seems more directly applicable.

In conclusion, a psychological theory would require considerable additions not provided for in Thagard's theory. The nature of these changes also recommends a conventional symbolic model rather than a connectionist one.

Assimilating evidence: The key to revision?

Michelle T. H. Chi
Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA 15260
Electronic mail: mlchk%orlcam@wms.cs.pittsburgh.edu

Thagard's theory of explanatory coherence has several exciting and profound applications in psychology. Two crucial but unresolved issues in psychology are: (1) How does conceptual change occur? and (2) What kind of "transition mechanism" accounts for these changes? These two questions most typically arise in the domains of learning and development. In learning, one manifestation of this issue concerns the transition from holding a naive theory of the physical world to holding a scientific or Newtonian view. In development, the issue concerns the transition from one stage (such as preoperational) to another, more advanced stage of thought (such as concrete operational). It is commonly thought that the shift from one kind of thought to another, either from preoperational to operational, or from pre-Newtonian to Newtonian, depends on the adoption of a set of interrelated beliefs. (This wholesale adoption is sometimes called radical restructuring.) The dilemma has always been trying to identify the "mechanisms" that enabled this transition to take place.

The most promising aspect of Thagard's theory is that it could potentially uncover precisely what factors can contribute to restructuring (or to conceptual change) without postulating an explicit mechanism that is responsible for the transition. That is, by implementing ECHO in a connectionist framework with parallel constraint satisfaction, the model has the capability of settling into a state "naturally," thereby achieving restructuring without identifying specific mechanisms for it. Thus, in some sense, ECHO has bypassed the problem of identifying the "transition mechanism" that has puzzled psychologists for decades. The implication of ECHO is that manipulating a few coherence relations in a piecemeal way might in fact produce dramatic shifts in one's theoretical orientation or frame of thought.

Although ECHO has this potential, what has ECHO accomplished so far? To understand what could have caused the transition (i.e., to understand what caused one theory to be more coherent than another), Thagard needs to model conceptual transitions directly. This is almost an impossible task in the historical context, somewhat less difficult in the developmental context, but perhaps feasible in a learning context. Thagard has attempted to model such a transition in the learning case by modeling the belief revisions that a student underwent in explaining the trajectories of projectiles while offering predictions and receiving feedback. In the data cited in Ranney and Thagard (1988), the shift exhibited by subject S.P.I. from a non-Newtonian to a Newtonian framework occurred primarily from encoding new evidence that either confirmed existing Newtonian hypotheses or contradicted existing non-Newtonian hypotheses. The advantage of this demonstration is that Ranney and Thagard could model the shift in conceptual change without postulating the formulation of new hypotheses, as was necessary in the historical cases (for example, Lavoisier had many new hypotheses that were not entertained by Stahl). This is fortunate because the mechanism by which new hypotheses are formulated is as yet little understood, as Thagard knows. What appears to have caused a shift from pre-Newtonian to Newtonian conceptions is modeled as the occurrence of new evidence either provided by the experimenter, or entertained by the student, evidence that either confirmed or contradicted the student's existing hypotheses (the student initially had both Newtonian and non-Newtonian hypotheses). Thus, in general, Thagard's applications of his theory to the learning domain, as well as to historical cases, point to two critical mechanisms needed for restructuring: the acquisition or formulation of new hypotheses and the encoding or entertaining of new evidence.

Unfortunately for ECHO's plausibility as a model of human performance, the majority of psychological evidence regarding conceptual change contradicts an implicit assumption underlying ECHO's analyses of Ranney's data. It is often found (in Piagetian research, for example) that confronting a child with evidence that contradicts the child's hypothesis usually does not lead to the child's rejection of that hypothesis. This suggests that the crux of the matter may not lie in the straightforward provision of new evidence, as modeled in Ranney and Thagard. Rather, the crucial insight for the subject is to realize that a particular hypothesis explains a particular piece of evidence. (This is the simplest case; I will avoid adding qualifications to all other cases, such as realizing that two hypotheses are contradictory, and so on.) Hence, what underlies theory revision may be precisely the willingness to adopt or reject a belief that a particular piece of evidence is explainable by a specific hypothesis. (This is currently built into ECHO as a given.) This willingness may in turn depend on the representation of a subject's current conception.

There is another issue that is relevant to psychology: What does ECHO's network of explanations represent? By modeling ECHO in a connectionist framework, Thagard is implying that the connectivity per se ought to inform the person whose mind the network embodies that a particular theory is more or less coherent than another. Presumably, if the network of Lavoisier's explanations is an accurate reflection of his memory, then it is not surprising that Lavoisier is convinced that his theory is the current one. This raises an interesting dilemma, however: Two contemporaneous theorists who hold opposing views would presumably know about each other's hypotheses as well as the evidence that each theory's hypotheses would explain. And yet, two contemporaneous theorists would not come to the same evaluation of their respective theories, as predicted from ECHO. This means that their representations must be different somehow. How they might differ can easily be seen in the arguments entered into by the theorists. Many of these arguments question the assumption that a particular hypothesis
explains a particular piece of evidence. This goes back to the previous point that the critical “insight” is the willingness to assimilate into one’s representation that a particular piece of evidence is explained by a particular hypothesis. As psychological evidence shows, one is unwilling to encode a piece of evidence and its interpretation if it conflicts with one’s existing hypotheses. Thus, we have cycled back to the original question of how exactly individuals revise their initial sets of beliefs in a significant way.

One other issue relating to ECHO’s feasibility as a human model concerns ECHO’s exhibition of apparently superhuman capabilities. In the behavioral decision-making literature, it has consistently been found that simple linear combinations of evidence are better at predicting outcomes such as success in graduate school (Dawes 1971) or the severity of Hodgkin’s disease (Einhorn 1972) than human experts (e.g., physicians in the case of diagnosing Hodgkin’s disease). The usual interpretation of these data is that humans excel at evaluating individual pieces of evidence with respect to a hypothesis but are extremely poor at integrating multiple pieces of evidence. The same superhuman reasoning ability may be exhibited by ECHO in that it can resolve two discrepant views given all their explanatory links, whereas humans only evaluate each individual explanation. The psychological community anxiously awaits further empirical tests to clarify these important issues.