Behavior Research Methods, Instruments, & Computers
1995, 27 (2), 239-243

8. COGNITIVE MODELS AND SIMULATIONS

Chaired by Michael Ranney, *University of California, Berkeley*

Protocol modeling, textual analysis, the bifurcation/bootstrapping method, and *Convince Me*: Computer-based techniques for studying beliefs and their revision

MICHAEL RANNEY and PATRICIA SCHANK
University of California, Berkeley, California

This paper traces a progression of four computer-based methods for studying and fostering both the structure and the on-line development of knowledge. Each empirical technique employs ECHO, a connectionist model that instantiates the theory of explanatory coherence (TEC). First, verbal protocols of subjects' reasonings were modeled post hoc. Next, ECHO predicted, a priori, subjects' text-based believability ratings. Later, the bifurcation/bootstrapping method was developed to elicit and account for individuals' background knowledge, while assessing intercoder reliability regarding ECHO simulations. Finally, *Convince Me*, our "reasoner's workbench," automated the explication both of subjects' knowledge bases and of their belief assessments; the *Convince Me* software permits contrasts between the model's predictions and subjects' proposition-wise evaluations. These experimental systems enhance our understanding of the relationships among—and determinant features regarding—hypotheses, evidence, and the arguments that incorporate them.

Human reasoning and argumentation represent some of the most vexing phenomena of cognitive psychology. Whether one is opining about O. J. Simpson's guilt or innocence at the local pub, or explaining new logic-puzzle data to colleagues, several difficulties arise. First, the extent of an individual's initial knowledge base is rarely clear: What does a person know, and what does a person *not* know? Determining the mechanisms by which people add to their knowledge is also difficult: How variable are individuals' "inference engines"? Finally, for a terminal corpus of beliefs, some propositions seem evidentiary, some seem more hypothetical, and all have varying confidence levels: How do we assess these features of a person's thinking?

Several attempts to account for the interrelationships, revisions, and/or structure of subjects' beliefs have been advanced—including schemata and mental models, "conceptual maps," discriminability analyses, and probabilistic belief networks (e.g., Austin & Shore, 1993; Bartlett, 1932; Carey, 1985; Chi, Feltovich, & Glaser, 1981; Gentner & Stevens, 1983; Pearl, 1988). These accounts have significant methodological or pragmatic limitations. For instance, concept maps, popularly used to contrast (relative) experts with novices, are commonly post hoc, seat-of-the-pants analyses by theorists who are "eyeballing" their data; even researchers who attempt to contrast explicitly the knowledge structures of two individuals or groups rarely (if ever) report intercoder reliability measures (see Ranney, in press). At another extreme, Bayesian-style probability networks have more rigor, but reasonably sized networks require many estimates of (e.g., conditional) probabilities that humans cannot consider, or have not pondered (Thagard, in press).

ECHO and the Theory of Explanatory Coherence

Between post hoc analyses of dubious reliability and rigorous but nonpragmatic techniques lie the theory of explanatory coherence (TEC) and its ECHO model (explanatory coherence by harmany [sic] optimization; Ranney & Thagard, 1988; Thagard, 1989). TEC includes roughly ten prominent principles of explanatory coherence

We thank the National Academy of Education, the Spencer Foundation, and the University of California's Committee on Research for supporting various aspects of this research. We especially extend thanks to Christine Diehl, Christopher Ritter, Christopher Hoadley, George Montoya, Kathryn Ann Dougery, Andrea Mosmann, and Jonathan Neff for their help on these various studies. Helpful conversations and comments from Brian Reiser, Paul Thagard, Carl Bereiter, Vimla Patel, Michelle Million, and other colleagues are similarly greatly appreciated. An earlier version of *Convince Me* is available for purchase as part of the 1994 *BioQUEST library* (Schank, Ranney, & Hoadley, 1994) by calling (301) 405-7600. More recent versions may become available for distribution; to inquire about such a possibility, and for correspondence regarding this article, write to M. Ranney, 4655 Tolman Hall, University of California, Berkeley, CA 94720-1670 (e-mail: ranney@cogsci.berkeley.edu).

(parsimony, contradiction, explanatory symmetry, data priority, propositional acceptance, system coherence, etc.; Ranney, in press; Thagard, 1989, 1992). ECHO implements TEC in a constraint-satisfying, connectionist program; beliefs become reified as localist representations—essentially sentence-sized statements about a particular controversy. The model passes activation—the "currency of believability"—among evidential and hypothetical propositions (nodes in a network), such that propositions that eventually exhibit high activation may be regarded as accepted, while propositions with low activation may be thought of as rejected. By itself, ECHO neither learns connection weights nor infers new propositional relationships; these are provided, depending upon the methodology employed (see below), either by default, by the experimenter, or by the subject.

Initial Efforts at Modeling Verbal Protocols with ECHO

Ranney and Thagard (1988) presented the first ECHO modeling of *on-line* human reasoning (in contrast to arguments extracted from scientific treatises; Thagard, 1989, 1992; cf. Miller & Read, 1991; Read & Marcus-Newhall, 1993). Using Ranney's (1987/1988) verbal protocols from subjects who were reasoning about ballistics, they modeled data from both rare and common conceptual difficulties. These subjects often achieved nontrivial Gestalt restructurings regarding inertia (e.g., in contrast to "impetus" perspectives; cf. Ranney, 1994b). For instance, one subject initially decided that objects that were dropped from a horizontally moving carrier (e.g., from a train window) would fall vertically, relative to the ground (i.e., the straight-down trajectory "S" in Ranney, 1994a); she later realized, upon considering motion at the apex of an arched trajectory, that such objects curve forward during their descents.

Belief revisions of this sort were modeled in ECHO by representing each of a subject's significant statements as either a piece of evidence or a hypothesis, with evidence afforded a measure of preferential treatment, in terms of activation (i.e., as specified by TEC's data-priority principle). The model also dictates that explanatorily linked statements excite each other, whereas contradictory and/or competitive propositions inhibit each other. Four main parameters (excitation, inhibition, data priority, and activational decay) were available for modulation, yet default values sufficiently modeled what the subjects believed or disbelieved during the various time segments that were analyzed. Thus, ECHO yielded activations that reasonably and temporally mimicked the changing assessments of beliefs by subjects as more information became available—information that resulted either from subjects' personal inferences or from external sources of feedback (Ranney & Thagard, 1988).

Predicting the Acceptance and Rejection of Beliefs Embedded in Text

Even the preceding, dynamic, post hoc simulations of protocols raised questions regarding the model's power, relative to the size of the data set (Ranney, in press). Hence, our later research (e.g., Ranney, Schank, Mosmann, & Montoya, 1993; Schank & Ranney, 1991) used ECHO *predictively*, such that activations generated a priori were contrasted with subjects' explicit ratings of the "believability" of propositions embedded in textual controversies that we provided as stimuli. The following modeled example (regarding the HIV virus) is a rich, ecologically realistic text from Christopher Ritter's (1991) work in our laboratory.

> A child who has tested positive for the presence of HIV (AIDS virus) wishes to enter a preschool. Are the other children in the school safe from becoming infected, or are they unsafe?
>
> On one hand, casual transmission of the infection may not be possible. 95% of all childhood HIV cases are known to have contracted the infection from their mothers at or before birth, or from receiving blood transfusions. The Surgeon General has determined that transmission of the HIV infection by casual contact is extremely unlikely. And no mother of an HIV-positive child (who has contracted HIV through transfusion) has become infected from her child. The unlikelihood of casual transmission would make it safe for the other children if the HIV-positive child were to attend the school.
>
> On the other hand, HIV transmission through casual contact may indeed be possible. 5% of pediatric HIV cases are of unknown origin. In a number of hospitals, AIDS patients are separated from other patients. And the virus has been demonstrated to be present in saliva and tears. Finally, the assumption that—under certain circumstances—all viruses can be casually transmitted suggests that casual transmission of HIV is possible. The likelihood of casual transmission of HIV would make it unsafe for uninfected children to attend school with the HIV-positive child.
>
> What do you think? (p. 35)

Such work was largely successful (modeling dynamic belief revisions by yielding activation-versus-rating correlations of up to .8), and allowed for the testing of several of the principles of TEC, but the subjects were clearly considering extratextual background knowledge in their deliberations (for instance, some parents mentioned that preschool children occasionally bite each other, making the Surgeon General's comments on casual transmission moot; on the other hand, some mentioned that excluding one's child would be discriminatory, and that it is preferable to risk contraction of the HIV virus). This led us to model even more explicitly subjects' reasoning and decision-making knowledge—and their relative coherence (cf. Ranney, 1994b)—as discussed below.

The Bifurcation/Bootstrapping Method Offers More Reliable Modeling

The advent of our resulting *bifurcation/bootstrapping method* represents a more formal foray into this knowledge-explication realm (see Ranney et al., 1993; Schank & Ranney, 1992). This method was developed to better model subjects' native (preexisting) knowledge, and to assess the intercoder reliability of researchers who model reasoning about such knowledge (cf. Ericsson & Simon, 1993): First, a subject's data (garnered from an intensive, semistructured interview) are bifurcated, with the person's explicit (numerical) believability ratings segmented out of a transcribed protocol, such that only the subject's beliefs (and the subject's assertions about the beliefs' interrelationships) remain. (In addition, more casual evaluative comments—such as "I *really* believe this"—are completely excised from the transcripts.) This "sanitized protocol" is then "bootstrapped" by several parallel human encoders, who generate and run their resulting encodings (as input) through ECHO, yielding activational predictions for every belief that the individual encoders extracted from the protocol. The activations, based upon the encoders' elucidations of a subject's propositions and propositional relations, are then blindly contrasted with the subject's sequestered believability ratings, providing an overall numerical correlational fit. Intercoder reliabilities arise by contrasting (1) correlations among activation sets and (2) the node-link topologies and structures yielded by the various encoders of the same protocol; see Figure 1 for an overview of this method. (Note that one need only replace "ECHO" in the figure with the name of a similarly formal, competing model of protocol-based belief assessment to see the generality of the bifurcation/bootstrapping method.)

Even though our initial results indicate both reasonably good data fitting and intercoder reliability, the bifurcation/bootstrapping method is fairly unwieldy. To obviate the arduousness of the technique, which requires an extremely vigilant and well-practiced experimenter, we now record comparable data (i.e., the structure, coherence, and evaluations of individuals' knowledge)—and even more data—in a more automated, yet rigorous, fashion.

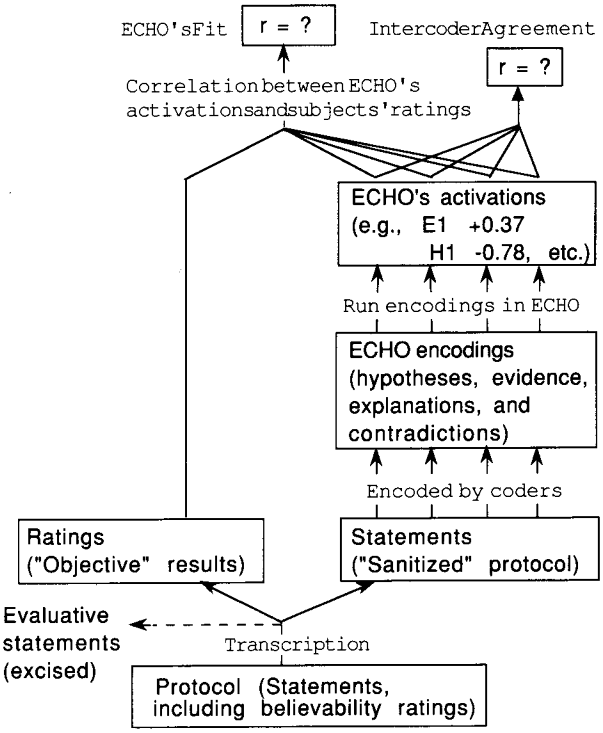
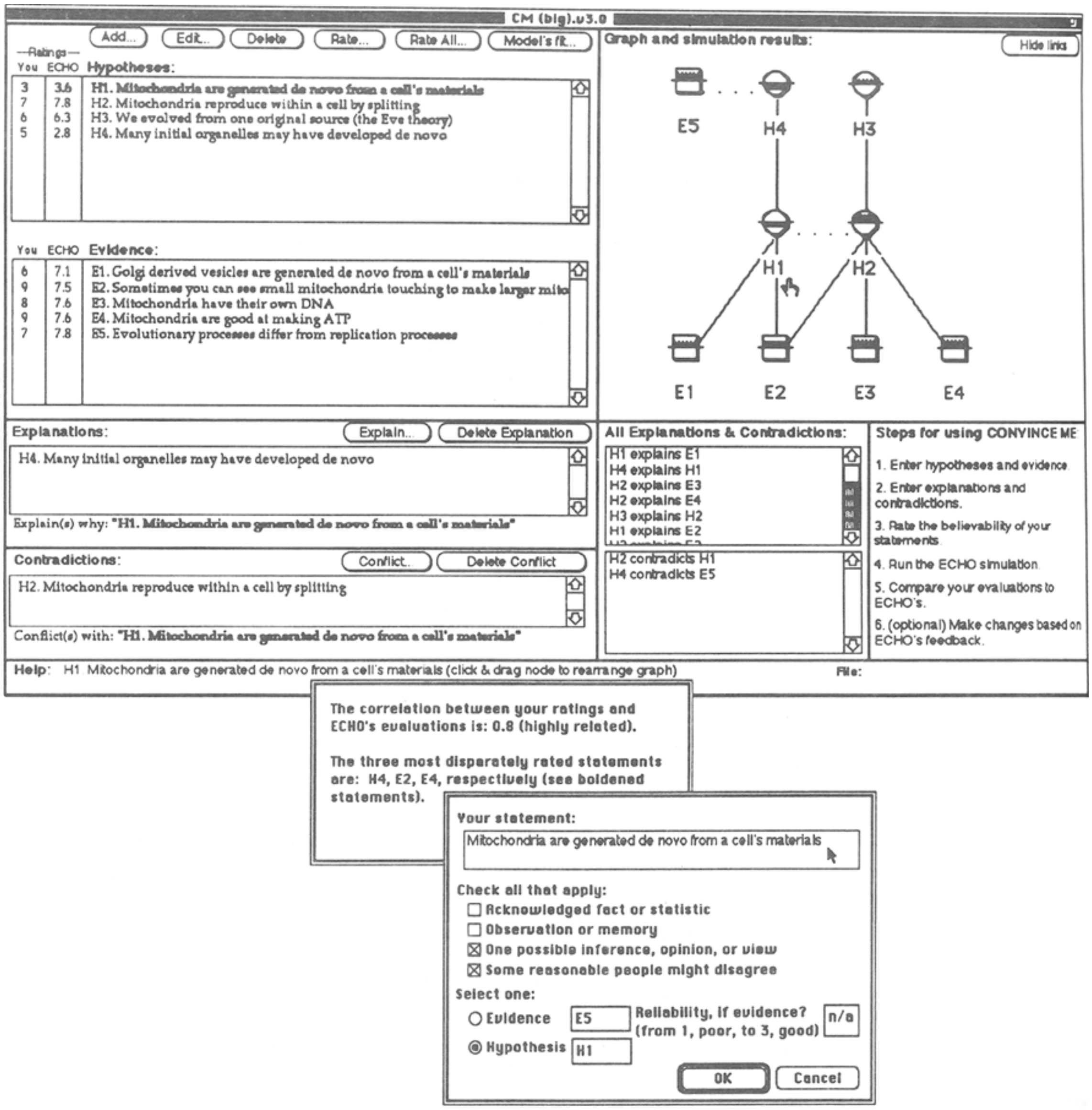

ECHO'sFit
r = ?
IntercoderAgreement
r = ?
CorrelationbetweenECHO's activationsandsubjects'ratings
ECHO's activations (e.g., E1 +0.37 H1 -0.78, etc.)
Run encodings in ECHO
ECHO encodings (hypotheses, evidence, explanations, and contradictions)
Encoded by coders
Ratings ("Objective" results)
Statements ("Sanitized" protocol)
Evaluative statements (excised)
Transcription
Protocol (Statements, including believability ratings)

Figure 1. A schematic summary of the bifurcation/bootstrapping method, as used with the ECHO model.

More Direct, Automated Knowledge Elicitation with *Convince Me*

Convince Me is a Macintosh program that captures both (1) a subject's "knowledge dump" of evidence and hypotheses—including their relationships and epistemic categorizations—and (2) believability ratings for these beliefs (Schank & Ranney, 1993; Schank, Ranney, & Hoadley, 1994). Rather than eliciting these in the relative maelstrom of an on-line interview/protocol session, *Convince Me* functions as a "reasoner's workbench" (Ranney, in press), upon which subjects explicate their beliefs about a controversy. The system guides students to cyclically (1) categorize their own propositions as either evidence or hypotheses; (2) indicate the reliability of their various pieces of evidence (a new feature used to modulate ECHO's data priority; cf. Schank & Ranney, 1991); (3) connect their propositions with both explanatory and contradictory/competitive links; and (4) rate each proposition's believability. After each (1–4) cycle, subjects can elicit feedback from ECHO to help improve the coherence of their knowledge bases (Ranney, Schank, Hoadley, & Neff, 1994). Figure 2 illustrates such a situation, in which the subject has already received feedback that is activation based (e.g., H1 = 3.6), correlational (e.g., r = .8), and proposition-wise (e.g., that H4, E2, and E4 represent the most discrepant pairs in the correlation). In Figure 2, the subject has subsequently chosen to examine Proposition H1 to consider modifying its wording or classification (see caption for more detail). The subjects are even permitted to alter the ECHO model if they feel that it doesn't "reason" as they do. Figure 3 indicates how a user may change the levels of "skepticism" (i.e., activational decay), and data "boost" (i.e., data priority), as well as the relative importance both of explanation (i.e., excitation) and of conflicts (i.e., inhibitory contradictions and competitions). However, the subjects rarely find it necessary, upon receiving feedback, to question ECHO's default parameters—they usually prefer first to explicate their arguments further.

Our empirical findings reinforce the claim that *Convince Me* is useful as a research tool with which people can progressively represent and evaluate more globally coherent bodies of information. Other assessments indicate that after they have finished with our system's training (e.g., during posttests), undergraduates distinguish better between hypotheses and evidence (Ranney, Schank, & Diehl, in press; Ranney et al., 1994), achieving the same discriminability levels between the two constructs as do experts in scientific reasoning (i.e., those who study such reasoning professionally; Ranney et al., 1994). Further, recent research indicates that both during and after employing *Convince Me* (e.g., amid exercises and later, on posttests) the subjects maintain an improved agreement between their believability ratings and their arguments' structures (Ranney et al., in press).



—Ratings—

You	ECHO	Hypotheses:
3	3.6	**H1. Mitochondria are generated de novo from a cell's materials**
7	7.8	H2. Mitochondria reproduce within a cell by splitting
6	6.3	H3. We evolved from one original source (the Eve theory)
5	2.8	H4. Many initial organelles may have developed de novo

You	ECHO	Evidence:
6	7.1	E1. Golgi derived vesicles are generated de novo from a cell's materials
9	7.5	E2. Sometimes you can see small mitochondria touching to make larger mito
8	7.6	E3. Mitochondria have their own DNA
9	7.6	E4. Mitochondria are good at making ATP
7	7.8	E5. Evolutionary processes differ from replication processes

Explanations:

H4. Many initial organelles may have developed de novo

Explain(s) why: "**H1. Mitochondria are generated de novo from a cell's materials**"

Contradictions:

H2. Mitochondria reproduce within a cell by splitting

Conflict(s) with: "**H1. Mitochondria are generated de novo from a cell's materials**"

All Explanations & Contradictions:

H1 explains E1
H4 explains H1
H2 explains E3
H2 explains E4
H3 explains H2
H1 explains E2

H2 contradicts H1
H4 contradicts E5

Steps for using CONVINCE ME

1. Enter hypotheses and evidence.
2. Enter explanations and contradictions.
3. Rate the believability of your statements.
4. Run the ECHO simulation.
5. Compare your evaluations to ECHO's.
6. (optional) Make changes based on ECHO's feedback.

Help: H1 Mitochondria are generated de novo from a cell's materials (click & drag node to rearrange graph)

File:

The correlation between your ratings and ECHO's evaluations is: 0.8 (highly related).

The three most disparately rated statements are: H4, E2, E4, respectively (see boldened statements).

Your statement:

Mitochondria are generated de novo from a cell's materials

Check all that apply:

□ Acknowledged fact or statistic
□ Observation or memory
☒ One possible inference, opinion, or view
☒ Some reasonable people might disagree

Select one:

○ Evidence E5 Reliability, if evidence? (from 1, poor, to 3, good) n/a
● Hypothesis H1

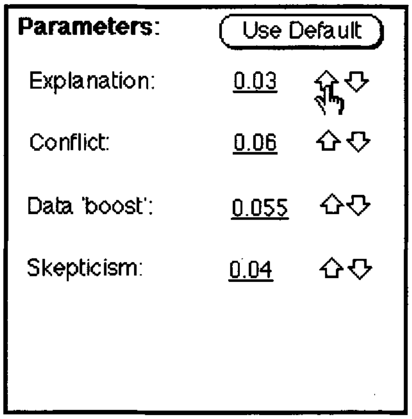
Figure 2. A subject reviews (at bottom) the characteristics of a belief (H1) upon receiving feedback (e.g., in the middle box) about a microbiological controversy. The subject's prior epistemic categorizations have resulted in the segmentation of beliefs into hypotheses and evidence, with associated believability ratings and contrasting, scaled ECHO values (top left). The diagram's node icons reflect the categorizations, while graphically providing the relative acceptance (and, for most arguments, the relative rejection as well) of the represented propositions (top right). The lower portions of the main window illustrate (1) some of H1's connectivity; (2) a listing of the network's connections; (3) a reminder of the basic steps for using *Convince Me*; and (4) a "help" facility that is keyed to the position of the "hand" cursor.

Further Prospects for Eliciting Knowledge and Modeling Arguments

Current prescriptive evaluation studies focus on the utility of diagrammatically representing *Convince Me*'s argument structures (see the networked "thermometer" icons in Figure 2), and the degree to which the software (and/or our associated curriculum) produces observed effects (e.g., Diehl, Ranney, & Schank, 1995; Ranney et al., in press). *Convince Me* may also be improved by incorporating more human processing limitations into its modeling. Ranney (in press) notes that human reasoning is rarely as globally coherent as that of a memorially infallible connectionist program, leading Hoadley, Ranney, and Schank (1994) to model subjects' data descriptively with "WanderECHO"—i.e., ECHO with a limited attentional capacity. Explicitly contrasting the feedback of ECHO with that of WanderECHO may make subjects more aware of localities in their own reasoning (e.g., in the ignoring of discordant information; cf. Chinn & Brewer, 1993).

Conclusions

We have sketched several computer-based methodologies for capturing subjects' reasoning, describing verbal-protocol data, and generating predictions of the plausi-



Parameters:

Explanation: 0.03

Conflict: 0.06

Data 'boost': 0.055

Skepticism: 0.04

Figure 3. A subject begins to modify ECHO's parameters in *Convince Me* (with the default parameter values shown).

bilities of individuals' problem- or text-based beliefs. Both our current results and our future focus largely center on our reasoner's workbench, *Convince Me*. The encouragement provided by our findings to date suggest that such systems can perhaps be to coherent reasoning what good word processors may be to writing—namely, handy recording tools that may enhance our thinking and even yield transfer to toolless ("posttest") environments.

REFERENCES

Austin, L. B., & Shore, B. M. (1993). Concept mapping of high and average achieving students and experts. *European Journal for High Ability*, **4**, 180-195.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, **5**, 121-152.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, **63**, 1-49.

Diehl, C., Ranney, M., & Schank, P. (1995, April). *Multiple representations for improving scientific thinking*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Gentner, D., & Stevens, A. L. (Eds.) (1983). *Mental models*. Hillsdale, NJ: Erlbaum.

Hoadley, C. M., Ranney, M., & Schank, P. (1994). WanderECHO: A connectionist simulation of limited coherence. In A. Rum & K. Eiselt (Eds.), *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 421-426). Hillsdale, NJ: Erlbaum.

Miller, L. C., & Read, S. J. (1991). On the coherence of mental models of persons and relationships: A knowledge structure approach. In F. Fincham & G. J. O. Fletcher (Eds.), *Cognition in close relationships* (pp. 69-99). Hillsdale, NJ: Erlbaum.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo: Morgan Kaufmann.

Ranney, M. (1988). Changing naive conceptions of motion. (Doctoral dissertation, University of Pittsburgh, Learning Research and Development Center, 1987.) *Dissertation Abstracts International*, **49**, 1975B.

Ranney, M. (1994a). Assessing and contrasting formal and informal/experiential understandings of trajectories. In G. H. Marks (Ed.), *Proceedings of the International Symposium on Mathematics/Science Education and Technology* (pp. 142-146), Charlottesville, VA: Association for the Advancement of Computing in Education.

Ranney, M. (1994b). Relative consistency and subjects' "theories" in domains such as naive physics: Common research difficulties illustrated by Cooke and Breedin. *Memory & Cognition*, **22**, 494-502.

Ranney, M. (in press). Explorations in explanatory coherence. In E. Bar-On, B. Eylon, & Z. Schertz (Eds.) *Designing intelligent learning environments: From cognitive analysis to computer implementation*. Norwood, NJ: Ablex.

Ranney, M., Schank, P., & Diehl, C. (in press). Competence and performance in critical reasoning: Reducing the gap by using *Convince Me*. *Psychology Teaching Review*.

Ranney, M., Schank, P., Hoadley, C., & Neff, J. (1994). "I know one when I see one": How (much) do hypotheses differ from evidence? In R. Fidel, C. Beghtol, B. H. Kwasnik, & P. J. Smith (Eds.), *Proceedings of the Fifth Annual American Society for Information Science SIG/CR Workshop on Classification Research* (pp. 139-156). [An updated version will appear in B. H. Kwasnik (Ed.) (in press), *Advances in classification research: Vol. 5*. (ASIS Monograph Series), Medford, NJ: Learned Information]

Ranney, M., Schank, P., Mosmann, A., & Montoya, G. (1993). Dynamic explanatory coherence with competing beliefs: Locally coherent reasoning and a proposed treatment. In T.-W. Chan (Ed.), *Proceedings of the International Conference on Computers in Education: Applications of Intelligent Computer Technologies* (pp. 101-106). Taipei, Taiwan, R.O.C.: The Artificial Intelligence in Education Society.

Ranney, M., & Thagard, P. (1988). Explanatory coherence and belief revision in naive physics. In V. L. Patel & G. J. Groen (Eds.), *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 426-432). Hillsdale, NJ: Erlbaum.

Read, S. J., & Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality & Social Psychology*, **65**, 429-447.

Ritter, C. (1991). *Thinking about ECHO*. Unpublished master's project, University of California, Berkeley.

Schank, P., & Ranney, M. (1991). An empirical investigation of the psychological fidelity of ECHO: Modeling an experimental study of explanatory coherence. In K. J. Hammond & D. Gentner (Eds.), *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* (pp. 892-897). Hillsdale, NJ: Erlbaum.

Schank, P., & Ranney, M. (1992). Assessing explanatory coherence: A new method for integrating verbal data with models of on-line belief revision. In J. K. Kruschke (Ed.), *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 599-604). Hillsdale, NJ: Erlbaum.

Schank, P., & Ranney, M. (1993). Can reasoning be taught? *Educator*, **7**, 16-21.

Schank, P., Ranney, M., & Hoadley, C. (1994). Convince Me [Computer program and manual]. In J. R. Jungck, V. Vaughan, J. N. Calley, N. S. Peterson, P. Soderberg, & J. Stewart (Eds.), *The BioQUEST Library*. College Park, MD: Academic Software Development Group, University of Maryland.

Thagard, P. (1989). Explanatory coherence. *Behavioral & Brain Sciences*, **12**, 435-502.

Thagard, P. (1992). *Conceptual revolutions*. Princeton, NJ: Princeton University Press.

Thagard, P. (in press). Probabilistic networks and explanatory coherence. In P. O'Rorke & J. Josephson (Eds.) *Automated abduction: Inference to the best explanation*. Menlo Park, CA: AAAI Press.

(Manuscript received November 18, 1994;
revision accepted for publication January 24, 1995.)