Modeling an Experimental Study of Explanatory Coherence

Patricia Schank and Michael Ranney

Education in Mathematics, Science and Technology
Graduate School of Education
University of California, Berkeley, CA 94720
schank@garnet.berkeley.edu, ranney@cogsci.berkeley.edu

Abstract

The problem of evaluating explanatory hypotheses is to choose the hypothesis or theory that best accounts for, or explains, the given evidence. Thagard (e.g., 1989) and Ranney (in press; Ranney & Thagard, 1988) describe a theory of *explanatory coherence* intended to account for a variety of explanatory evaluations; this theory has been implemented in a connectionist computer model, ECHO. In this study, we examine three questions regarding the relationship between human explanatory reasoning and ECHO's explanatory evaluations: Does ECHO predict subjects' evaluations of interrelated propositions? Are local temporal order differences (not explicitly modeled by ECHO) important to the subjects? Does ECHO predict subjects' inflectional reasoning? We found that subjects often entertain competing hypotheses as nonexclusive and presume an implied backing for certain (superordinate) hypotheses. These tendencies were modeled in ECHO by assigning a fraction of *data priority* (usually reserved for evidence) to the superordinate hypotheses. In sum, the ECHO model helps to interpret subjects' reasoning patterns, and shows continued potential for simulating explanatory coherence processes.

Introduction

The assessment of explanatory hypotheses involves choosing the hypothesis or theory that best accounts for given evidence. The theory of *explanatory coherence* (e.g., Thagard 1989) accounts for a variety of explanatory evaluations. In the theory, relations of local explanatory coherence and incoherence between *propositions*, interpreted as hypotheses or evidence, are established via several principles. In general, propositions cohere with propositions they explain, propositions they are explained by, or propositions they are analogous with; they incohere with competing or contradictory propositions. The *acceptability* of a proposition depends on the acceptability of the propositions to which it is related, as its acceptability generally increases with the coherence of the sub-system of propositions in which it is embedded. Propositions interpreted as evidence generally accumulate acceptability more rapidly than do hypotheses, though the acceptability of a datum may decrease if it is mostly coherent with propositions that are unacceptable.

The theory of explanatory coherence has been implemented in a connectionist computer model called ECHO. In ECHO, propositions are represented by nodes in a network that send and receive *activation* (the measure of acceptability) to and from related propositions. Hypothesis evaluation is treated as the parallel satisfaction of multiple constraints derived from explanatory and contradictory relations. Once an ECHO network has settled, propositions with high activations are considered accepted (Thagard 1989, in press2). ECHO differs, however, from networks in which node-node link weights are adjusted on each cycle (see Rumelhart & McClelland 1986).

What is the relationship between human explanatory reasoning and explanatory evaluation as implemented in ECHO? ECHO has been used in the past to model juror reasoning (Thagard 1989) in which explanatory coherence plays a crucial role (Pennington & Hastie 1988), to understand mental models of social interactions and relationships (Miller & Read in press), and to model scientific reasoning (Thagard 1989, in press1). Ranney and Thagard (1988), in the first application of ECHO to modeling experimental data, simulated (*ex post facto*) changes in subjects' beliefs and their conceptions of physical motion (Ranney 1987/1988; cf. Nersessian 1989). Ranney (in press) suggests that if ECHO continues to successfully model human reasoning, it might be useful as a tool for teaching coherent argumentation; however, speculation about the relationship between ECHO's descriptive power and prescriptive potential requires that we address several empirical questions about the psychological fidelity of ECHO. We describe here an empirical study that examined ECHO's ability to model explanatory evaluations of textually embedded propositions (Schank 1991).

The ECHO Model

Coherence involves a relation between two propositions that "hold together," while incoherent propositions resist holding together. Relations of coherence (and incoherence) among propositions are established via the following principles (Thagard in press1; Ranney & Thagard 1988): (1) *Symmetry*: Coherence and incoherence are symmetric relations. (2) *Explanation*: Hypotheses that together explain a proposition cohere with each other, as well as with the explained proposition. (3) *Simplicity*: The coherence of an explanation is inversely related to the number of co-hypotheses used in the explanation. (4) *Analogy*: Similar hypotheses that explain similar evidence cohere. (None of our networks involved analogies.) (5) *Data Priority*: Results of observations, such as evidence and acknowledged facts, have a degree of acceptability on their own. (6) *Contradiction*: Contradictory hypotheses incohere. (7) *Competition*: Competing hypotheses (which explain the same evidence or hypotheses but are not themselves explanatorily related; see Harman 1989) incohere. (This principle did not affect the ECHO networks used in this study since such competitions were explicitly made contradictory; the competition principle is used to *generate* incoherence links.) (8) *Acceptability*: The acceptability (or believability) of a proposition depends on its coherence within the system of propositions in which it is embedded. A proposition's acceptability increases as its coherence with other acceptable propositions increases. (A proposition's acceptability is measured by its activation value, ranging from -1, complete rejection, to 1, complete acceptance.)

Hypothesis evaluation in ECHO is treated as the satisfaction of multiple constraints derived from the explanatory relations, and several parameters provide degrees of freedom (Thagard 1989, in press2). The number of parameters used in computer models often leads to questions: How arbitrary are the simulation findings? How stable is the model under a given set of parameters (i.e., does one set work over a wide range of tasks)? Should some parameters be regarded as variant or invariant over tasks? Thousands of simulations have been done to determine the impact of parameter settings on ECHO's performance, and the values of four parameters—the *excitation*, *inhibition*, *decay*, and *data excitation* rates—appear most critical (Thagard 1989; Ranney & Thagard 1988). However, the criteria used to decide the default values of these parameters have previously been practical rather than empirical. Ranney and Thagard (1988) as well as this study provide the first empirical backing for assigned parameter values.

The excitation value determines the weights on links between cohering propositions. The inhibition value determines the weights on links between incohering propositions. The data excitation value specifies the weight on links between data (usually evidence) and the "special evidence unit" (SEU, a unit with activation clamped at 1.0) to implement principle 4. The decay value specifies the percentage of the (absolute) activation that a proposition loses at each cycle. Thagard (1989) describes interesting psychological interpretations of these parameters. The *tolerance* of the system is the absolute value of the ratio of excitation to inhibition; highly tolerant systems may "believe" several competing hypotheses, while systems with low tolerance tend to accept a single theory. Decay is the *skepticism* of the system; as decay increases, asymptotic activation values of propositions will be compressed toward zero (although the simulations will generally not be qualitatively different). Skepticism applies directly to all propositions, while tolerance is most directly relevant to contradictory or competing propositions.

Given declared input propositions and relations between them, ECHO creates units to represent the propositions, and sets up the following links (with additive link weights): (1) excitatory links between propositions (including co-hypotheses) that are explanatorily or analogously related; (2) inhibitory links between contradictory and (nonevidential) competing propositions; and (3) excitatory links between evidence and the SEU (with the data excitation weight). (The data priority of particular evidence may also be specified separately.) If there are unexplained data, the decay rate is increased appropriately (i.e., unexplained evidence raises skepticism). Unit activations are updated in cycles until the network settles and the change in all units is asymptotic. At each cycle, the activation of a particular unit u_j is determined by the decay rate and the net input to the unit (the weighted sum of the activation of its neighboring units, where a weight is the common link value, w_{ij}), and is updated using the following equation (cf. Rumelhart & McClelland 1986):

$$u_j(t+1) = u_j(t)\,(1\text{-}decay) + \begin{cases} net_j\,(max\text{-}u_j(t)) & \text{if } net_j > 0 \\ net_j\,(u_j(t)\text{-}min) & otherwise \end{cases}$$

where $net_j = \sum_i w_{ij}\, u_j(t)$

Evaluating ECHO

We ask three questions about the relationship between human reasoning and ECHO: How well does ECHO predict differences in subjects' evaluations of different texts? Are local temporal order differences (not explicitly accounted for by ECHO) important to the subjects? How well does ECHO predict inflectional reasoning, in which new information presented in texts (cf. Kintsch 1988) yields significant changes in subjects' beliefs (Ranney in press)?

To investigate these questions we conducted three contrasts between ECHO simulations and subjects' believabilities. Each contrast involved three stages: (1) ECHO simulations were run on given systems of propositions, (2) subjects read textual embodiments of the same propositional system and rated how strongly they believed each proposition, and (3) ECHO's ability to model the subjects' believability ratings was assessed.

Contrast One: Differential Predictions. Given textual propositions, how well do subjects' beliefs in the propositions correlate with ECHO's activations for these propositions? For this contrast, two ECHO topologies were used, the second of which includes an extra, critical piece of evidence that qualitatively changes the simulation results. If two groups of subjects are given texts reflecting these topologies (with one group given the extra piece of information), does ECHO predict the differences among the groups' beliefs in the underlying propositions? The major empirical question is: Are subjects' believability ratings reasonably consistent with ECHO's activations?

Contrast Two: Temporal Order Effects. Does the order of presentation of theories in a text affect the subjects' believability ratings? Given that all propositions are generally encoded during the same network cycle, ECHO activations are unaffected by the temporal order of propositional encoding. If two groups of subjects are given texts reflecting a single network topology, with the second group given a text that provides the propositions in a different order, will their beliefs in the propositions differ? If so, the finding could be used to suggest modifications to ECHO.

Contrast Three: Dynamic Modeling. With its use of data priority, ECHO is somewhat empirically biased. Are people as empirically driven as ECHO? Once subjects are familiar with a text, does adding an additional piece of information—one that produces a "Gestalt switch" in ECHO—induce a Gestalt switch in the subjects? To assess this contrast, an ECHO topology was allowed to run and settle, and then an additional piece of evidence—one that yielded a theory shift reflected by inflected activation trajectories—was introduced (see Ranney & Thagard 1988). Do subjects also shift from accepting one theory to accepting another upon the introduction of the new evidence? Will adding an additional hypothesis, one that strengthens ECHO's acceptance of the newly acceptable theory, further strengthen subjects' beliefs in the theory? Perhaps the new hypothesis will shift subjects' beliefs toward the new theory if the additional evidence did not already do so. The absence of a change when evidence is introduced may suggest that people are more hypothesis-driven than is ECHO. The absence of any shift at all (or a shift only when both the datum and the hypothesis are introduced) may suggest that people are more predictively intransigent (Ranney 1987/1988) than ECHO, and serve as a basis for re-evaluating ECHO's data priority aspect.

Method and Design

The texts used in this study reflect portions of the topology shown in Figure 1, which represents two conflicting theories, and lends itself to the questions in the above contrasts. In this topology, hypothesis H1 explains evidence E1 and E2, and is in turn explained by hypothesis H0. Hypothesis H2 explains evidence E2, E3, and E4, and is explained by hypothesis H3. Subgraphs of this topology, shown in Figures 2 and 3, are also used for the contrasts of interest.

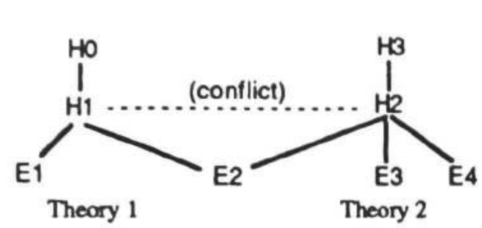


Figure 1. Topology of two conflicting explanatory theories; dashes indicate theoretical conflict.

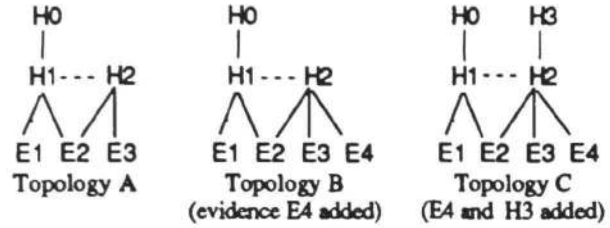


Figure 2. Topologies A, B, and C. (Topologies A and B were used for the differential predictions contrast. Topologies A, B, and C were used for the dynamic modeling contrast.)

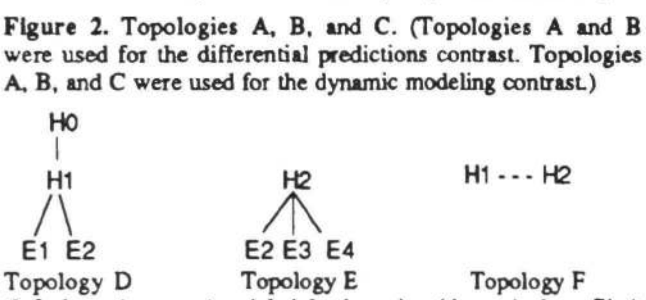


Figure 3. Topologies D, E, and F (used to assess temporal ordering effects).

Table 1. Four sequences x four texts.

Topology Sequences			
S1	S2	S3	S4
A	D	E	B
<rate>	D+F	E+F	<rate>
B	D+F+E	E+F+D	
<rate>	<rate>	<rate>	
C			
<rate>			
Decontextualized Texts			
T1	T2	T3	T4
linguistics	physics	wine naming	medical

Texts involving fictional theoretical controversies were generated to reduce the chance of subjects bringing considerable extraneous information into their analyses. Since dimensions of context are difficult to assess and control, we devised four texts in different domains. Subjects read the texts (given on separate sheets of paper) under no time limit, and then rated (on a separate questionnaire sheet; see Table 1) how strongly they believed the text's statements.

Materials

The four texts were designed to be fairly context-free, while alternating between science and humanities domains. Four different topological sequences, S1 through S4, were used to constrain the structures of these texts (see Table 1). Sixteen text-sequence combinations resulted (4 sequences X 4 texts, e.g., T1-S1, T1-S2...). The physics text below represents topology C; its ECHO encoding is shown in Table 2.

"Sue and Amy are trying to identify an atomic particle:

Sue and Amy find that it has a charge magnitude of two electrons, high energy, a positive spin, and a baryon number equal to zero. Sue says that physicists think that ziptons are rather common, which explains her belief that the particle is a zipton. Sue also notes that if one assumes that the particle is a zipton, that explains why the particle has a charge magnitude of two electrons and why it has high energy.

Amy disagrees with Sue. Amy believes that the particle is a blinkon, which contradicts Sue's belief that the particle is a zipton. Amy says that physicists think that blinkons are easy to detect, which explains her belief that the particle is a blinkon. Amy also notes that if one assumes that the particle is a blinkon, that explains why the particle has high energy, why it has a positive spin, and why it has a baryon number equal to zero.

What do you think?"

Table 2. Encoding of the physics text, topology C.

Evidence:	
proposition E1 "the particle has the charge of two electrons"	
proposition E2 "the particle has high energy"	
proposition E3 "the particle has a baryon number of zero"	
proposition E4 "the particle has positive spin"	
Hypotheses:	
proposition H0 "physicists think ziptons are rather common"	
proposition H1 "the particle is a zipton"	
proposition H2 "the particle is a blinkon"	
proposition H3 "physicists think blinkons are easy to detect"	
Explanations:	
H0 explains H1	
H1 explains E1	**Contradictions:**
H1 explains E2	H1 contradicts H2
H3 explains H2	**Data priority for evidence:**
H2 explains E2	data E1 E2 E3 E4
H2 explains E3	
H2 explains E4	

Subjects and Procedure

To control for presentation orders, all 24 possible topological sequence combinations were used, and both forward (T1, T2, T3, T4) and backward (T4, T3, T2, T1) text orders were used. In total, 48 (4! x 2) subjects participated. Subjects had various academic and occupational backgrounds, and ranged in age from 18 to 60 years.

Each experimental session consisted of four parts. In each part, the subject was given a text reflecting one of the text-sequence combinations, never receiving the same text or the same topology sequence twice. For each text-sequence combination, a subject was asked to read the text and then rate the believability of its propositions. Ratings were on a scale from 1, "completely unbelievable," to 7, "completely believable" (4 was unlabeled). When the subjects were finished rating all four text-sequence combinations, they were asked (via audio-taped retrospective protocols), to explain how they came up with their believability assignments—in order to determine if they brought extraneous information into their evaluations.

The ECHO simulations were initially run with parameter settings used by Ranney and Thagard (1988), the first application of ECHO to modeling experimental data. Simulations were subsequently run with slightly adjusted parameter settings, as discussed below.

Contrast One, Topology A vs. Topology B: Differential Predictions. Does ECHO differentially predict subjects' responses to different texts? ECHO activations for topologies A and B were compared in a pairwise manner with subjects' ratings on propositions from Sequences 1 (topology A) and 4 (topology B, including its extra piece of evidence, E4). (Subjects were given additional text and questionnaires beyond topology A in Sequence 1; see Contrast Three below and Table 1.) The added datum significantly changes ECHO's simulation: Without the extra evidence E4 and the H2-E4 explanatory link, hypothesis H1 receives higher activation than H2 in ECHO; with the extra evidence and link, the reverse is true.

Contrast Two, Sequence 2 vs. Sequence 3: Temporal Order Effects. How sensitive are subjects to the temporal order of propositions? Running ECHO on topology B, hypothesis H2 receives a higher activation than H1, given that all propositions are encoded during the same network cycle. For this contrast, subjects were asked to read three portions of text (presented on separate sheets) reflecting topology B, in which the order of sets of propositions in the texts were varied. Topology B can be decomposed into three subgraphs, representing a "left" theory, a partial "right" theory, and a theoretical conflict (topologies D, E, and F in Figure 3). For Sequence 2, subjects were given three successive texts: the first reflecting the left theory, the next reflecting the left theory and theoretical conflict, and the last reflecting the left and the partial right theories as well as the theoretical conflict (topology order: D, D+F, D+F+E). For Sequence 3, subjects were also given three successive texts, with the order of the left and partial right theories switched (E, E+F, E+F+D).

Contrast Three, Sequence 1: Dynamic Modeling. How well does ECHO dynamically model inflectional reasoning? In ECHO, hypothesis H1 receives a higher activation than H2 in topology A (see Figures 2 and 5). After the network settles, adding evidence E4 and the explanatory link H2-E4 causes the activation of proposition H1 to decrease and H2 to increase, surpassing H1, thus simulating a mild Gestalt shift. Adding the hypothesis H3 and the explanatory link H3–H2 causes H2's activation to further increase and H1's to further decrease. For this contrast, all subjects were given a text structurally reflecting topology A, followed by a believability rating questionnaire. They were then given a text reflecting topology B, followed by a new questionnaire. Finally, the subjects were given a text reflecting topology C, followed by a final questionnaire. The subjects' ratings were then compared to activations from ECHO simulations during which propositions E4 and H3 (and their associated links) were added dynamically (i.e., in later network cycles, after successive settlings). If subjects are as empirically driven as is ECHO, they should "switch" from favoring H1 to favoring H2 when E4 is added, and certainly with the addition of H3.

Results and Discussion

Interviews with the subjects revealed that they adopted some kind of implicit *presumed backing* (i.e., other evidence or beliefs) for the superordinate hypotheses H0 and H3. The ECHO simulations were adjusted to take this into account. Thus, before describing the fit between the subjects' ratings and ECHO's activations, we first discuss the subjects' believability ratings and comments to show how presumed backing affected their ratings.

Subjects' Believability Ratings

The means of subjects' believability ratings (for propositions in topologies A, B, and C are shown in Table 3. Subjects rated the superordinate H0 higher than its explained H1, and the superordinate H3 higher than its explained H2. Subjects also tended to view H1 and H2 as nonexclusive hypotheses (in spite of the explicitly stated conflicts), particularly on the medical text.

Table 3. Mean ratings for all appropriate propositions.

	Topology (within Sequence)					
Prop	S1a	S1b	S1c	S2b	S3b	S4b
E1	6.50	6.56	6.60	6.52	6.58	6.65
E2	6.50	6.58	6.58	6.31	6.67	6.65
E3	6.52	6.58	6.56	6.19	6.50	6.50
E4	•	6.42	6.58	6.19	6.46	6.63
H0	4.81	4.73	4.69	4.65	4.60	4.88
H1	3.58	3.27	3.35	3.46	3.29	3.41
H2	3.48	4.12	4.29	3.85	3.98	4.08
H3	•	•	4.73	•	•	•

Presumed Backing. Some subjects justified an implicit backing of the superordinate hypotheses by indicating that the word "think" (in the texts) is used for conversational politeness; i.e., people use the word "think" when they really *know* something. Despite the attempt to generate decontextualized texts, subjects (not surprisingly; Ritter 1991) also considered other knowledge when reasoning about the texts, and this other knowledge likely contributed to their presumed backing of H0 and H3. For instance, one subject hinged his ratings of the wine text on his belief that there are few types of wine, unlike atomic particles, of which there "are many." In contrast, another subject held that there are *not* many types of atomic particles! Although subjects generally gave high ratings to evidence, these ratings still did not boost their beliefs in hypotheses H1 and H2 above their superordinates.

Text Type and Text Order Effects. As expected, an analysis of variance on subjects' ratings (on the Topology C texts) revealed that effects of text type and text order were not significant ($p > .05$). There was, however, a (nonsignificant) trend for the superordinate H0 to resist being overcome by H3 with increasing experience with the successive texts. (Nb. Many of this paper's inferential tests are based on standard statistical distributional assumptions that remain to be confirmed for this particular paradigm.)

ECHO Simulations

In contrast to the subjects, *a priori* ECHO activations of the superordinates H0 and H3 were lower than the hypotheses they explained (H1 and H2) across a wide range of parameter settings. Also, ECHO's activations of the conflicting hypotheses H1 and H2 were slightly more divergent than the subjects' ratings of these hypotheses. This suggested that the apparent differences between subjects' ratings and ECHO's activations were due to a presumed backing of hypotheses H0 and H3 and the tendency of subjects to view H1 and H2 as nonexclusive.

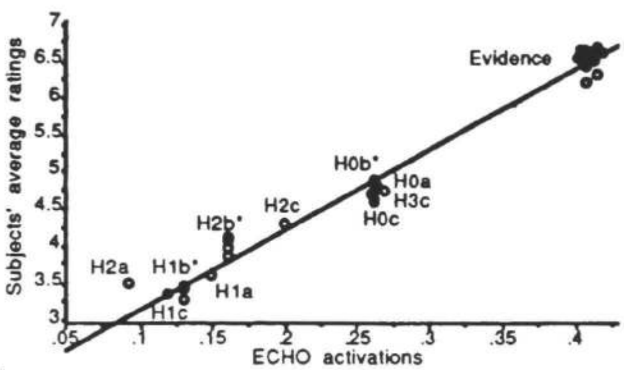
To model subjects' presumed backing, H0 and H3 were given partial data priority by linking them to the SEU with a weight half of that given to evidence. This was encoded by adding these semi-hypotheses to the data statement:

data E1 E2 E3 E4 (H0 .5) (H3 .5)

The parameter values used for four simulations (see below) are shown in Table 4 ("R&T" refers to values used in Ranney & Thagard 1988).

Table 4. ECHO parameter settings on simulations.

Parameters	—no backing—		—backing—	
	Run 1 (R&T)	Run 2 (adjusted)	Run 3 (R&T)	Run 4 (adjusted)
excitation	.03	.005	.03	.03
inhibition	-.06	-.06	-.06	-.06
decay	.04	.15	.04	.15
data excitation	.055	.10	.055	.10



*multiple occurances, from ratings over Sequences S1 to S4

Figure 4. Relation between *mean* believability ratings and ECHO's activations with presumed backing.

Table 5. Point-by-point ranked correlations between ECHO activations and subjects' belief ratings.

	—no backing—		—backing—	
	Run 1	Run 2	Run 3	Run 4
SEQUENCE	(R&T)	(adjusted)	(R&T)	(adjusted)
S1a	.71	.67	.71	.78
S1b	.71	.76	.78	.78
S1c	.56	.68	.73	.76
S2b	.63	.70	.73	.73
S3b	.73	.79	.80	.80
S4b	.71	.78	.80	.80
TEXT				
linguistics	.66	.72	.76	.77
physics	.61	.67	.73	.76
wine	.68	.76	.78	.77
medical	.72	.79	.80	.81
OVERALL	**.67**	**.74**	**.77**	**.78**

Comparing Subjects and ECHO

Figure 4 shows the relation between the subjects' *mean* ratings and ECHO's predictions when presumed backing is taken into account. Point-by-point correlations between individual subjects' ranked ratings and ECHO's ranked activations were also calculated. Table 5 lists these correlations for different sequences, different texts, and overall.

As shown in Figure 4, subjects' ratings are modeled best when presumed backing is represented in ECHO, and when the decay and data excitation values are both slightly raised, muting the differences between H1 and H2 (see Tables 4 and 5, Run 4). As indicated in Table 5, the overall correlation between the subjects' responses and ECHO is high (r_{ranked}=.78).

Contrast One: Differential Predictions. An analysis of variance revealed that across topologies A and B, subjects' ratings of hypothesis H1 did not significantly change ($p>.1$; see Table 3). However, as predicted by ECHO, subjects' ratings of hypothesis H2 rose significantly across topologies A and B ($p<.05$) with the introduction of the evidence E4 (which H2 explains). On topology A (S1a), subjects' ratings of hypothesis H1 were not significantly higher than those of H2 ($p>.1$). However, on topology B (S4), subjects' ratings of hypothesis H2 were significantly higher than those of H1 ($p<.05$), as ECHO predicts. When presumed backing for hypothesis H0 is represented in ECHO, the relation between the subjects' ratings and activations on topologies A and B is particularly strong because the subjects' higher ratings of H0, compared to H1, are captured. With presumed backing, the point-by-point correlation between the subjects' averages and ECHO is high on both topology A (r_{ranked}=.71-78), and topology B (r_{ranked}=.80). (See Table 5, Runs 3 and 4, S1a and S4b.)

Contrast Two: Temporal Order Effects. No differences between subjects' ratings of the propositions in Sequences S2b and S3b (and even S1b and S4b) approached significance (see Table 3, S2b and S3b). With presumed backing, the correlation between the subjects' averages and ECHO is high for both S2 and S3 (Table 5, Runs 3 and 4). Nonsignificant hints of subjects' intransigence were shown by ECHO's slightly lower correlation with subjects' ratings for S2 (in which the "losing" theory is given first), and by a smaller variance in subjects' ratings for S3 (in which the "winning" theory is given first). Still, as shown in Table 3, the subjects' mean ratings are almost identical on topology B for S1 and S4 (S1b and S4b), even with different subjects and an intervening rating in S1, indicating few "prediction-freezing" effects. Although one subject remarked after reading the S1 (dynamic) version of the medical text that "the more information that comes up, the less I believe them 'cause they didn't know it before," the local temporal order differences—which ECHO does not automatically take into account—were not generally found to be important.

Contrast Three: Dynamic Modeling. Across topologies A (S1a), B (S1b), and C (S1c), subjects' ratings of hypothesis H1 did not significantly change ($p>.1$; see Table 3). However, subjects' ratings of hypothesis H2 rose significantly across topologies ($p<.05$). As predicted by ECHO, the introduction of the new evidence E4 (in S1b) accounted for most of subjects' increased believability in H2 ($p<.05$), while the introduction of the superordinate H3 (in S1c) had little extra effect on subjects' believability of H2 ($p>.1$). On topology A, the subjects' ratings of hypothesis H1 were not quite significantly higher than H2; however, on topologies B and C, subjects' ratings of hypothesis H2 were significantly higher than H1 (both tests p<.05). Thus, both the subjects and ECHO show inflectional reasoning when the new evidence of topology B is introduced. The relation between the subjects' ratings and the ECHO activations is once again particularly strong when presumed backing of the superordinate hypotheses is represented in ECHO. Figure 5 illustrates the inflectional reasoning captured by ECHO (e.g., for H1 and H2). Table 5 shows the high correlations between the subjects' ratings and ECHO's activations for topologies A, B, and C (rows S1a, S1b, and S1c).

Like ECHO, the subjects "switched" and rated H2 higher than H1 when the new evidence was introduced (see Table 6). With presumed backing, the correlation between subjects and ECHO is also high when H3 (of topology C) is added (see Table 5, row S1c).

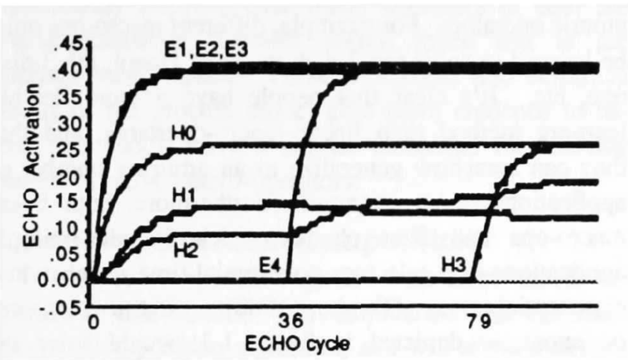


Figure 5. ECHO activation trajectories, Topology C, S1 (evidence E4 and hypothesis H3 added dynamically), Run 4.

Table 6. Subjects' average believability ratings on S1, suggesting inflectional reasoning.

	Time 1: S1a	Time 2: S1b (E4 added)	Time 3: S1c (E4 & H3 added)
H1	3.58	3.27	3.35
H2	3.48	4.12[ab]	4.29[ab]

[a]significantly higher than H1 (p<.05).
[b]significantly higher than H2, Time 1 (p<.05).

Summary and Conclusion

In all contrasts, subjects' believability ratings were modeled well by ECHO. Subjects' tendencies to presume a kind of unstated backing behind the superordinate hypotheses were modeled in ECHO by giving a measure of data priority to these hypotheses and by slightly raising decay (skepticism) and data excitation. The differences in subjects' responses to texts reflecting topologies A and B were captured by ECHO, particularly when the presumed backing was represented. Also, local temporal order differences did not significantly affect subjects' beliefs, and since ECHO does not automatically take such order differences into account, the results do not suggest necessary changes to the model. Finally, ECHO captured subjects' inflectional reasoning—again, particularly well when the presumed backing was invoked. The theory of explanatory coherence and its ECHO model were useful in that they allowed us to interpret patterns in a subject's reasoning process fairly concisely. ECHO's ability to model the subjects' propositional evaluations well, even without the adjustments described above, suggests that ECHO shows continued promise for predicting human explanatory reasoning.

The subjects indicated that they considered some extraneous knowledge when reasoning about the texts, and this implicit knowledge was modeled in ECHO by giving certain hypotheses a fraction of data priority. However, this suggests that the subjects' representations of the situations were not completely captured by the representation encoded into ECHO. Thus, we are now investigating ECHO's ability to model individual subjects' beliefs about physical motion when they make their "implied backings" explicit.

Acknowledgements

We thank Paul Thagard, Boaz Keysar, Pete Pirolli, Marcia Linn, Mark Wilson, Jeff Schank, Chris Ritter, Peggy Carlock, Ed Lagache, Chris Borton, Mimi Recker, Steve Adams, and Michelle Million for their helpful comments.

References

Harman, G. 1989. Competition for evidential support. *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Erlbaum.

Kintsch, W. 1988. The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*, 163-182.

Miller, L. C., & Read, S. J. in press. On the coherence of mental models of persons and relationships: A knowledge structure approach. To appear in G.J.O. Fletcher and F. Fincham Eds. *Cognition in close relationships*. Hillsdale, NJ: Erlbaum.

Nersessian, N. J. 1989. Conceptual change in science and in science education. *Synthese, 80*, 163-183.

Pennington, N. & Hastie, R. 1988. Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 521-533.

Ranney, M. 1988. Changing naive conceptions of motion. Ph.D. diss., University of Pittsburgh, Learning Research and Development Center, 1987. *Dissertation Abstracts International, 49*, 1975B.

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

Ranney, M. & Thagard, P. 1988. Explanatory coherence and belief revision in naive physics. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*, 426-432. Hillsdale, NJ: Erlbaum.

Ritter, C. 1991. *Thinking about ECHO*. M.A. thesis, University of California, Berkeley, School of Education.

Rumelhart, D. E., McClelland, J. R. & the PDP Research Group. 1986. *Parallel distributed processing: Explorations in the microstructure of cognition*, vols. 1 and 2. Cambridge, MA: MIT Press.

Schank, P. 1991. *An Empirical Investigation of the Psychological Fidelity of ECHO*. M.A. thesis, University of California, Berkeley, School of Education.

Thagard, P. 1989. Explanatory coherence. *Behavioral and Brain Sciences, 12*, 435-502.

Thagard, P. in press1. The dinosaur debate: Explanatory coherence and the problem of competing Hypotheses. In J. Pollock and R. Cummins (Eds.), *Philosophy and AI: Essays at the interface*. Cambridge, MA: MIT Press/Bradford Books.

Thagard, P. in press2. *Conceptual revolutions*. Draft, June, 1990. Princeton, NJ: Princeton University Press.