[In the (1994) *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, pp. 421-426, Hillsdale, NJ: Erlbaum.]

WanderECHO: A Connectionist Simulation of Limited Coherence

Christopher M. Hoadley
Graduate Group in Science and
Mathematics Education (SESAME)
University of California at Berkeley
Berkeley, CA 94720
tophe@garnet.Berkeley.EDU

Michael Ranney
Institute of Cognitive Studies and
Graduate School of Education
University of California at Berkeley
Berkeley, CA 94720
ranney@cogsci.Berkeley.EDU

Patricia Schank
Education in Mathematics, Science,
and Technology (EMST)
University of California at Berkeley
Berkeley, CA 94720
schank@garnet.Berkeley.EDU

Abstract

The Theory of Explanatory Coherence, or TEC, (Ranney & Thagard, 1988; Thagard, 1989, 1992) and ECHO, a connectionist implementation of TEC, attempt to model human reasoning about evidence and hypotheses. The ECHO model is based on the simultaneous satisfaction of multiple constraints. This yields predicted activations ("believabilities") for propositions, which are based on the propositions' evidential status, their explanatory relationships, and their contradictory relationships. While ECHO has been demonstrated to usefully model human reasoning, it does not model processing limitations on the maintenance of coherence. WanderECHO is a variation on the ECHO model that attempts to simulate attentional and memorial limitations with a stochastic updating algorithm that is based on a traveling focus of attention. Several variants of the WanderECHO simulation were applied to Schank and Ranney's (1991) data, and were found to generally simulate subjects' mean believability ratings better than standard ECHO.

Introduction

The problem of limited rationality is a pernicious one in artificial intelligence and psychology. Little is known about how and when people maintain consistency and coherence in their beliefs. Consider the following example from Ranney (in press-a):

> We may often have the impression that people are remarkably adept at maintaining coherent mental models and representations. For example, if you had just told a friend that the city marathon is to be held today, you would not be terribly surprised if she quickly concludes, "Well, in that case, I should move my car." Indeed, her inference would be a testament to the everyday power of human coherence-seeking...
>
> By way of contrast...You tell a visiting friend what your mother's maiden name was (e.g., "Smith"), and he correctly, but unexpectedly, concludes, "Well, in that case I've got to move my car."... For instance, perhaps the friend has remembered [a] that your crazy uncle, who lives in the neighborhood, dumps trash on autos that park in front of his house, [b] that you mentioned that you come from a very small extended family, and [c] that he just parked in front of a house emblazoned with the sign, "The Smiths." A coherent inference or two, e.g. that your mother and uncle had the same name, leads to his conclusion.

The first example seems a likely example of inference; the second seems highly unlikely. We do not expect people to draw all possible inferences among pieces of knowledge, nor is it a computationally practical solution—even for a complex artificial intelligence system with a large knowledge base.

This paper investigates WanderECHO, a computer model designed to simulate aspects of limited rationality in humans. The model stochastically simulates aspects of attention and short-term memory limitations. It is based upon the ECHO computer model of coherence, which has proven useful as a model of human reasoning (Miller & Read, 1991; Ranney, Schank, Mosmann, & Montoya, 1993; Read & Marcus-Newhall, 1993; Read & Miller, 1993; Schank & Ranney, 1991, 1992). Both ECHO and WanderECHO are localist connectionist systems that employ simulations to embody the simultaneous constraint satisfaction described in the Theory of Explanatory Coherence (TEC; Ranney & Thagard, 1988; Thagard, 1989, 1992). However, WanderECHO uses a stochastic updating algorithm that is based on the link weights between nodes. In this paper, the model is compared empirically to both human data and ECHO simulations, and found to be an improvement over the standard ECHO simulation.

The ECHO Model

The Theory of Explanatory Coherence (e.g., Ranney & Thagard, 1988; Thagard, 1989) treats coherence as a multiple constraint satisfaction problem, where coherence is a continuous quantity to be optimized. Constraints are not rigid, but "soft" (cf. Smolensky, 1988). The theory is based upon the following principles: symmetry, explanation, simplicity, analogy, data priority, contradiction, acceptability (believability), and system coherence. The theory describes three related constructs: coherence, system coherence, and believability. Believability is a property of a single proposition that describes the degree to which the proposition is believed or accepted as true. Coherence is a property of a pair of propositions that describes the degree to which they are mutually compatible. (This is presumably generally determined by "background" or "world" knowledge.) System coherence is a property of a system of propositions. It is proposed that people try to maximize system coherence among their beliefs.

ECHO is a symbolic connectionist embodiment of the principles of TEC. It is based on a localist representation in which nodes represent propositions, and links between nodes represent coherence relations. Activations for nodes represent believability, weights on links between nodes represent a measure of coherence, and system coherence is the inverse of the system energy function. Therefore, a network with high energy has low coherence and vice versa. The network optimizes system coherence by settling to a low energy state. (In principle, these low energy states can be only local minima instead of the desired global minimum. By starting with near-zero positive activations, 0.01 for all units, this danger is reduced.) Data priority is represented by creating links between the evidential propositions and the *special evidence unit*, or SEU, whose activation is fixed at 1.0.

In most psychological studies carried out with ECHO, only three link weights among propositions are generally used: zero, a default inhibitory weight, or a default excitatory weight (but cf. Schank & Ranney, 1993). Likewise, only two link weights are used for links to the SEU: an excitatory weight, or zero. Default values for each of these parameters are given below in Table 1. These parameters have been used in various studies of ECHO (Ranney & Thagard, 1988; Schank & Ranney, 1991, 1992) Other values for these parameters have also been used, however. In one of their variations, Schank and Ranney (1991) adjusted certain link weights to account for the phenomenon of "presumed backing," in which certain propositions were treated as partially evidential, due to links to background knowledge that was not part of the information presented to or discussed with subjects.

Table 1: Some ECHO parameter settings

Parameter	Default	Schank & Ranney '91
Excitatory/Explanatory link weight	0.03	0.005
Inhibitory/Contradictory link weight	-0.06	-0.06
Data priority—SEU link weight	0.055	0.10
Decay	0.04	0.15
Stop value	0.01	0.01

ECHO also has a decay parameter, which moderates the activation of each node at each update cycle. A fixed percentage of the absolute value of the activation is sapped at each cycle; thus, nodes that do not have some external source of activation will, over time, tend towards zero.

The WanderECHO Model

Attention and Memory

As the example from the introduction demonstrates, people are not globally coherent all the time, nor do we expect this. Humans can only do limited coherence seeking. Can this be simulated? In particular, humans have a limited amount of time and limited memory with which to maintain coherence among beliefs. Attention and the order of processing also manifest effects. Primacy and recency biases are well-documented effects in both the history of science (Thagard & Nowak, 1990) and cognitive science (Ranney, Schank, Mosmann & Montoya, 1993; Stich, 1990). Work by Schank and Ranney and others (e.g., Ranney, 1987/1988; Ranney, et al., 1993; Ranney & Thagard, 1988; Schank, 1990; Schank & Ranney, 1992; Schank, Ranney, Hoadley, Diehl, & Neff, in press) demonstrates that belief judgments both change in observable time and are sensitive to subjects' attention.

ECHO is flawed in its representation of human thought in at least two aspects: computational power and memory limitations. A general problem is that the model does not account for the limits of human computational power. The ECHO simulation will run on a network of almost unlimited size and will continue updating all activations until the entire network settles. However, humans can not maintain coherence on even moderately large sets of beliefs without external assistance (Ranney, 1987/1988, in press-a, in-press-b). A fascinating example is provided by Tweney (1991), as he provides an account of the elaborate record keeping and cross-referencing techniques employed by Faraday in integrating information from his career of electricity experiments. Faraday constructed notebooks full of information and spent a great deal of time studying and cross-indexing this information in an effort to build a coherent view. Clearly, coherence maintenance on such a large scale is neither easy nor characteristic of our own mental processes, and in Faraday's case it could not have been performed at all without extensive external aids. Large problems are not simply processed until entirely done.

A more specific problem with ECHO is that it does not take into account short-term memory limitations. Computationally, this may be seen as a limit on the space aspect of space-time tradeoffs in calculation. This makes ECHO somewhat unrealistic: in an ECHO simulation, every node, no matter how many there are, is updated on every cycle. While the size of a chunk is certainly ambiguous, it seems safe to say that updating seven nodes at a time would be near the upper limit for conscious processing (cf. Miller, 1956). The only way for humans to exceed their short-term memory limitations is to do what modern serial computers do when working on problems too large to fit in main memory: swap information in and out for parts of the problem (cf. Ericsson, Chase, & Faloon, 1980; Simon, 1974).

The WanderECHO Algorithm

WanderECHO is a variation on ECHO that tries to take some of these considerations into account. First, WanderECHO simulates a limited focus of attention and does not demand massively parallel execution. Second, WanderECHO has a local stopping criterion, and does not require calculation of the energy change of the entire system in order to determine whether or not to stop; it will satisfice, rather than optimize as ECHO does. The model operates stochastically and can produce a range of different belief activations for the same network topology on different runs.

Rather than updating every unit in the network at every cycle, the WanderECHO algorithm updates one unit at a time. At any given cycle, a single node is updated; this node is the model's "focus of attention." The first node to be updated is chosen randomly. The next node to be updated is chosen probabilistically based on link weights. Given the current node i, the probability that a neighboring node x will be updated next is given by:

$$P(update_x) = \frac{|w_{ix}|}{\sum_{j \neq i} |w_{ij}|}$$

where j varies over all nodes not equal to i.

This concept of wandering attention has its roots in activation-spreading theories of memory. Since the strength of a link between two nodes provides a loose measure of how related the propositions are, this system allows attention to follow chains of related ideas.

Since the system is not calculating new activations for every node in the network at each cycle, it is neither desirable nor possible to use a global stopping criterion such as the total change in activations over the network. Rather than base the stopping criterion on only the change in activation in the current node, a time-averaging system is used. This is since a node may be in a state of local equilibrium at a given cycle, even though the entirety of the network may be decidedly unsettled. To maintain a rough equivalency with parameter settings for the ECHO model, the absolute values of the changes in activation over the previous n individual node updates are summed and compared to the stop criterion, where n is the number of nodes in the network.[1] Thus, in ECHO, one cycle of an eight-node network would update each of the eight units simultaneously, and compare the total change to the stop criterion. In a WanderECHO simulation of an eight-unit network, though, the change over a group of eight individual node updates would be summed and compared. These eight updates might occur over all eight nodes or it might occur over as few as two nodes. This makes WanderECHO likely to stop earlier than ECHO for equivalent criterial stopping values. If all eight nodes are successively updated, the criterion is the same as that for ECHO, and if fewer nodes are successively visited, less change is likely to happen.

Another important difference implemented for the WanderECHO simulation is its wider range of possibilities for starting activations. Under ECHO, nodes are all started at an equal, fixed activation, usually near zero. This makes sense for finding the overall settling point for the network; by starting the system with small values, wild oscillations that could occur with extreme starting activations are avoided, and spurious local energy minima that could result from unusual configurations of activations are avoided. However, in WanderECHO, small starting activations almost guarantee premature stopping; since local coherence everywhere will initially be relatively high except between units with data priority and the special evidence unit, WanderECHO paths that do not visit evidence nodes often enough will come to an inappropriate, early halt. The software provides options for different starting states: the user may ask that all nodes be started with a fixed value as with ECHO, or they may request random positive activations. (Random zero-centered starting activations were also possible, but were generally not successful, were not applicable to systems with activations ranging from 0 to 1, and are not reported here.)

In the limit, WanderECHO and ECHO will calculate similar values. If the WanderECHO stopping criterion is stringent enough, WanderECHO will continue to update nodes one by one until an energy minimum for the system is reached. WanderECHO spends more time updating highly connected nodes, so less connected nodes will probably take longer to settle; this means that the overall simulation will take more processing cycles than does ECHO to reach identical values. Neglecting the effect of stochastic updating on decay, the energy function for each simulation is the same.

At this point it may be helpful to take a step back and examine the reasons for creating the WanderECHO algorithm. One could argue that varying parameters or changing update rules is simply tinkering with connectionism's gory details, which seems unnecessary given the successes of the connectionist paradigm. However, ECHO is not primarily an application of connectionism but rather an instantiation of a principled philosophical theory of coherence. ECHO networks represent high-level features of cognition, not microscopic detail. WanderECHO considers high-level propositions one at a time, just as a person might. It is not a prescription for neural net simulations in general.

Empirical assessment of WanderECHO

The Human Data

This study used human data collected by Schank and Ranney (1991). Forty-eight subjects were presented with some subset of twelve texts. Each text described a fictional story, involving hypotheses and evidence, in one of four domains for which prior knowledge was unlikely to significantly contribute. The texts, when diagrammed as a network of hypotheses and evidence, produced one of three topologies (an example of which is shown below in Figure 1). In this way, it was hoped that subjects would represent the situation with the same propositions and links used to run the simulations. Subjects were asked to rate their relative

[1] This would prove unrealistic for large networks for the same reasons that updating large networks in cycles proves unrealistic. A fixed number of updates to average over is probably more realistic. However, in fairly small networks of the sort we employ here, the difference is minimal.

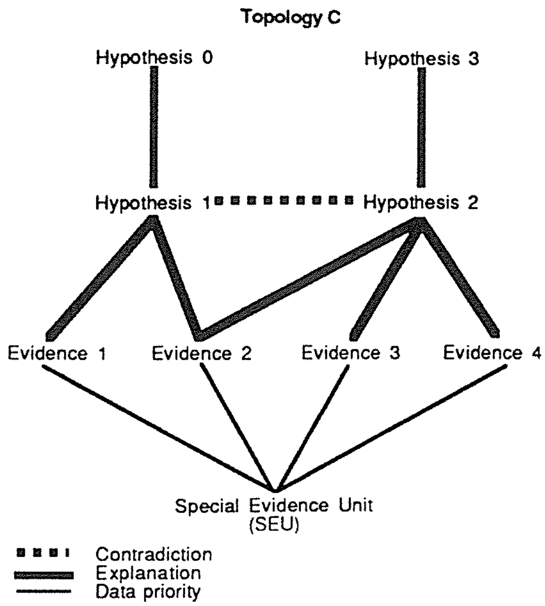

Figure 1: One of the network topologies.

belief in each of the propositions contained in the text, on a scale from 1 (completely unbelievable) to 7 (completely believable).

Comparison Methodology

Since the WanderECHO simulation is stochastic, comparing human results to an individual run of the model is inappropriate.[2] Because of this, the model was run 200 times, and then output activations were averaged across the 200 different runs. These average activations were then correlated with human believability ratings. To assay some of the properties of the new model, various combinations of parameter settings were used. Some of these parameters are from the original ECHO model, while some are specific to the WanderECHO algorithm.

Initially, over 25 different settings of parameters were used with WanderECHO to see if quantities such as link weights, starting values, or decay should be varied from ECHO settings. The original ECHO settings appeared adequate with two exceptions: starting node activation, and stop threshold. WanderECHO was more effective when starting activations were non zero (see below). For stop threshold, using a less stringent stopping value of 0.1 (compared to ECHO's 0.01) improved WanderECHO's accuracy.

For the final comparisons, three orthogonal parameters were varied in the WanderECHO model. The first parameter to be varied was starting activation. For half of the final trials, activations were started at 0.5. For the other half of the trials, activations were started at random positive values. Second, the nodes' minimum activation was varied; in half of the trials, activations ranged from -1 to 1, while in the other trials, activations ranged from 0 to 1. Since decay brings activations toward zero, in the first case, decay was pulling activations towards neutrality, and in the second case it was pulling them toward "disbelief." The third parameter to be varied was the update algorithm. In half of the trials, WanderECHO's focus of attention was allowed to wander through the Special Evidence Unit (SEU). (Note that even so, the SEU's activation was never updated, since it has an activation of 1.0 by definition.) In the other half, focus of attention was only allowed to wander across explanatory or inhibitory links. By allowing the focus of attention to travel through the SEU, all evidential nodes were effectively linked, and one data node after another might be updated. Without this "SEU wanderthrough," data nodes would typically only be visited by way of the hypotheses they explained. (In ECHO networks, evidence is not typically linked to other evidence.)

[2]Some form of individual-to-simulated-run matching would be required. A WanderECHO simulation might be constrained to a particular update path based on individual protocols, such as those in Schank & Ranney (1992).

Results

Table 2 shows the Pearson correlations between the average human responses and the ten comparison conditions: eight WanderECHO and two regular ECHO. One topology had eight propositions, one had seven, and one had six, yielding twenty-one data points. All the WanderECHO correlations were significantly greater than zero at the $p < .01$ level using the Dunn test. Most of the correlations were numerically greater than that of the "best-parameters" ECHO simulation (Schank & Ranney, 1991), and all modeled the data significantly better than ECHO with default parameters.

Table 2: Correlations between mean (WanderECHO and regular ECHO) activations and mean human ratings for each node in each topology.

WanderECHO, Minimum Activation = -1.0

	Wander path	
Start Values	SEU not permitted	SEU permitted
Random positive	r =0.931	r =0.965
0.5	r =0.953	r =0.961

WanderECHO, Minimum Activation = 0.0

	Wander path	
Start Values	SEU not permitted	SEU permitted
Random positive	r =0.914	r =0.963
0.5	r =0.945	r =0.962

Regular ECHO

ECHO, standard parameters	r =0.724
ECHO, "best" post-hoc parameters (Schank & Ranney, 1991)	r =0.941

For one topology, Table 3 shows the averages, by proposition, for the human data, the best and standard or default ECHO simulations, and one WanderECHO simulation of that topology (random positive start values, minimum activation of -1.0, SEU paths permitted). Figure 2 plots these results normalized as Z-scores. Note that the ECHO "best" simulation actually pins all hypotheses near neutral (i.e., zero activation). Also note that an even higher correspondence can be observed for the WanderECHO model if one presumes the existence of a ceiling effect on the human rating scale; in the simulations, Evidence proposition 2 produces a somewhat higher activation than the other evidence units, but since most people rated all the evidence units at or near the maximum of 7, a slightly higher belief rating for one piece of evidence over the others is difficult to measure.

Table 3: Averaged values for Topology B (subject ratings, ECHO output activations, and average WanderECHO output activations).

Prop.	Human	ECHO default	ECHO best	Wander-ECHO[3]
E1	6.578	0.708	0.396	0.550
E2	6.522	0.749	0.396	0.594
E3	6.443	0.715	0.396	0.567
E4	6.422	0.715	0.396	0.560
H0	4.714	0.333	0.001	0.455
H1	3.365	0.510	0.014	0.388
H2	4.016	0.605	0.032	0.416

Discussion and Conclusions

Stochastic updating is one viable way to simulate attention and limited processing in ECHO. While the current data set is too small to be conclusive, the results are encouraging regarding WanderECHO's prospects for becoming a useful simulation of limited coherence. However, introducing randomness into the model is both blessing and curse. On the positive side, the model may be able to simulate distributions of belief sets across a population. Indeed, WanderECHO's output distributions appear roughly similar to human data. However, testing this hypothesis requires a better understanding of the correspondence between activations and behaviors, such as ratings reported on a Likert scale (cf. Schank & Ranney, 1991; Schank et al., in press). As suggested above, ceiling or floor effects on the Likert scale may limit the ability to assess the correspondence between the model and human ratings. Furthermore, one would like to be able to attribute individual differences to parameters in the model, such as starting activations or update paths. Still, by adding more parameters to the model we also increase the risk of post-hoc curve fitting and of models with little predictive value. In this case, the additions of simulated attention and variable initial belief states represented a calculated risk based on the usefulness of these constructs in psychology. This choice must be validated by further operationalizing these two concepts, measuring them, and constraining WanderECHO to reproduce them faithfully.

[3]Minimum activation -1.0, SEU permitted wander path, random positive starting activations.

Acknowledgments

We thank the members of the Reasoning Group for their helpful suggestions and comments. We also especially thank Jonathan Neff, Christine Diehl, Terry Regier, Rafael Granados, and Michelle Million for their feedback.

References

Ericsson, K.A., Chase, W.G., & Faloon, S. (1980). Acquisition of a memory skill. *Science, 208*, 1181-1182.

Miller, G.A. (1956). The magical number seven, plus or minus two. *Psychological Review, 63*, 81-97.

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

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. (in press-a) 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. (in press-b). Relative consistency and subjects' "theories" in domains such as naive physics: Common research difficulties illustrated by Cooke and Breedin. *Memory & Cognition.*

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).

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.

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

Read, S.J., & Miller, L.C. (1993). Explanatory coherence in the construction of mental models of others. *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, 836-841. Hillsdale, NJ: Erlbaum.

Schank, P. (1990). *An empirical investigation of the psychological fidelity of ECHO*. First year project report, University of California, Graduate School of Education, Berkeley, CA.

Schank, P., & Ranney, M. (1991). The psychological fidelity of ECHO: Modeling an experimental study of explanatory coherence. *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, 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. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, 599-604. Hillsdale, NJ: Erlbaum.

Schank, P., & Ranney, M. (1993). Can reasoning be taught? *Educator, 7* (1), 16-21. [Special issue on Cognitive Science and Education.]

Schank, P., Ranney, M., Hoadley, C., Diehl, C., & Neff, J. (in press). A reasoner's workbench for improving scientific thinking: Assessing *Convince Me*. *Proceedings of the 1994 International Symposium on Mathematics/Science Education and Technology*. AACE.

Simon, H.A. (1974). How big is a chunk? *Science, 183*, 482-488.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11*, 1-74.

Stich, S. (1990) Rationality. In D. Osherson and E. Smith (Eds.), *An invitation to cognitive science: Thinking* (pp. 173-196). Cambridge, MA: MIT Press.

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

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

Thagard, P. & Nowak, G. (1990). The conceptual structure of geologic revolution. In J. Schrager & P. Langley (Ed.), P. Langley (Series Ed.), *Computational models of scientific discovery and theory formation*. San Mateo, CA: Morgan Kaufmann Publishers.

Tweney, R. D. (1991). Serial and parallel processing in scientific discovery. In R. N. Giere (Ed.), *Cognitive models of science* (pp. 77-88). Minneapolis: University of Minnesota Press.

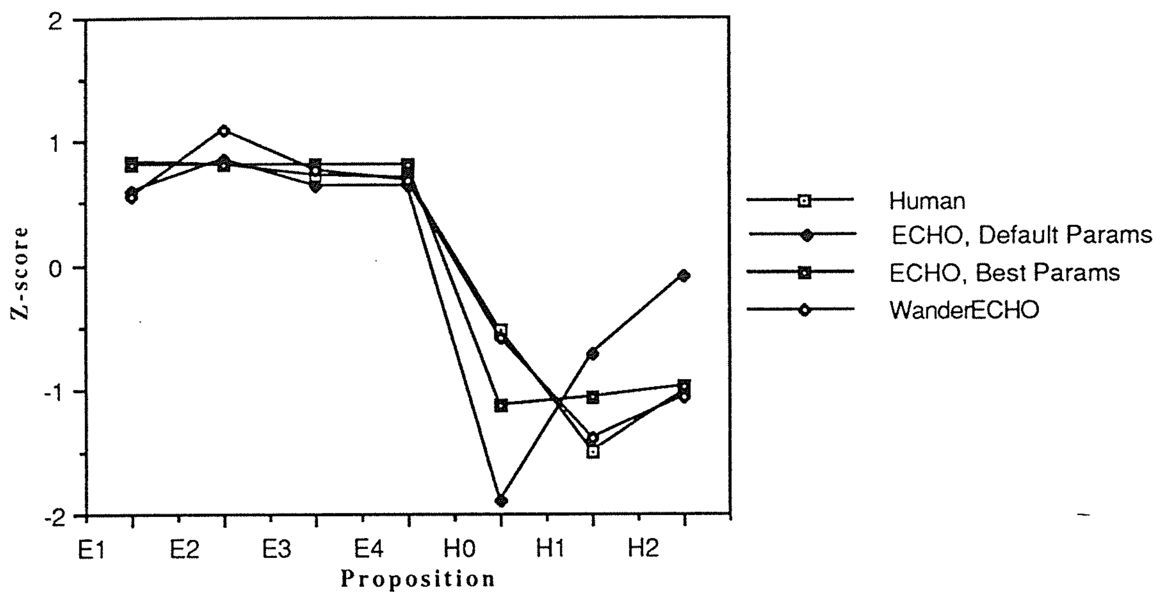

Figure 2: Normalized average values for activations/ratings from ECHO, WanderECHO, and humans—for Topology B.