Improved Reasoning with *Convince Me*

Patricia Schank and *Michael Ranney*
Mathematics, Science, and Technology Division, Graduate School of Education
University of California, Berkeley, CA 94720
(510) 654-8931
schank@garnet.berkeley.edu, ranney@cogsci.berkeley.edu

ABSTRACT

This paper describes *Convince Me*, a tool for generating and analyzing arguments. Results indicate that the system makes people better reasoners while they employ it, and yields transfer to situations unsupported by the software.

KEYWORDS: Connectionism, reasoning, instruction.

INTRODUCTION

Over the past few years, we've developed a progression of computer-based methods for studying beliefs. Each technique employs the *Theory of Explanatory Coherence* (TEC), and its connectionist implementation, ECHO [3, 5, 7]. We have found that ECHO usefully predicts people's evaluations of the hypotheses and evidence from their structured arguments [5, 6]. Based on these studies, we developed a TEC-based "reasoner's workbench" computer program—*Convince Me*—and a curriculum for teaching "scientific" reasoning [3, 4]. While other formal systems exist for the analysis and generation of arguments [e.g., 1], it seems that no other that includes a computational, theory-based model that yields predictions about the plausibility of an argument's many propositions. Five hours of training with the *Convince Me* software and curriculum made novices behave more like "scientific reasoning" experts; they more strongly (a) discriminated evidence from hypothesis, (b) doubted statements they rated as more hypothetical, and (c) associated believability with evidence-likeness [5]. While the distinguishing characteristics of data and theory are still vague (even for experts), our system lends sophistication to novices' discriminative criteria. But how much of these gains are due to the software versus the curriculum? Do they transfer to unsupported practice? This study assesses the software's effectiveness by contrasting people's performance under two conditions: doing exercises with *Convince Me* versus on paper, given identical written tests, instructions, and curriculum.

Figure 1 shows a person's argument regarding abortion in *Convince Me*. People can use the program to enter their ideas (bottom right, Figure 1), indicate which ideas explain and contradict which others, rate how strongly they believe each notion, and run an ECHO simulation to see which ones their argument helped to support, reject, or leave neutral—from the simulation's point of view. They can also ask *Convince Me* to report (a) a "model's fit" correlation between their ratings and ECHO's activation—with lexical labels (e.g., "mildly opposed", "highly related"), and (b) which (three) pairs of ratings differ the most (middle right, Figure 1). Based on this feedback, users can modify their ratings and/or argument (e.g., focusing on the statements for which they and ECHO most "disagree"), or even adjust the ECHO model to better simulate their thinking. However, users rarely opt for the latter—they usually prefer to further explicate their arguments first.

One measure of the utility of the software involves how well people's beliefs are in accord with their argument's structure. Prior work [e.g., 5, 6] suggests that an increased correlation between a person's and ECHO's "believability" values would indicate that the person's argument better reflects his or her beliefs—whether garnered from *Convince Me*, or via paper and pencil (and simulated later by an experimenter). Another interesting measure relates to the kinds of changes people employ when making revisions. For instance, do *Convince Me* users make more changes to their argument structure, or do they just superficially revise their belief ratings in whatever direction will give them a better correlational fit with ECHO?

METHOD

Twenty University of California, Berkeley undergraduates participated. They had various backgrounds, but essentially none in logic or the philosophy of science. The subjects completed a pre-test, three curriculum units on scientific reasoning, integrative exercises, a post-test, and an exit questionnaire [see 5]. Half completed the integrated exercises using *Convince Me* (the "Convince Me Group"); the other half used paper and pencil (the "Written Group"). Both groups were given the same prompts to generate arguments, give ratings, and make any revisions. Analyses revealed that there were no significant differences between the two groups in age, year in school, SAT scores, or total session hours.

RESULTS

Our findings replicate the essential results of [5] regarding hypotheses and evidence, and suggest that *Convince Me*'s knowledge-eliciting interface and simulation-driven feedback are critical for subjects' learning.

During the exercises, *Convince Me* users' beliefs were more in accord with the structures of their arguments, as evidenced by belief-activation correlations ($p<.05$; Figure 2). The software also yielded tranfer: Belief-activation correlations for *Convince Me* users (a) did not significantly dip during the post-test, when they did not have access to the software, and (b) were higher than those for both their own pre-test, and the Written Group's post-test ($p<.05$). In contrast, correlations for the Written subjects rose less during the exercises, and their post-test was nonsignificantly higher than their pre-test performance.

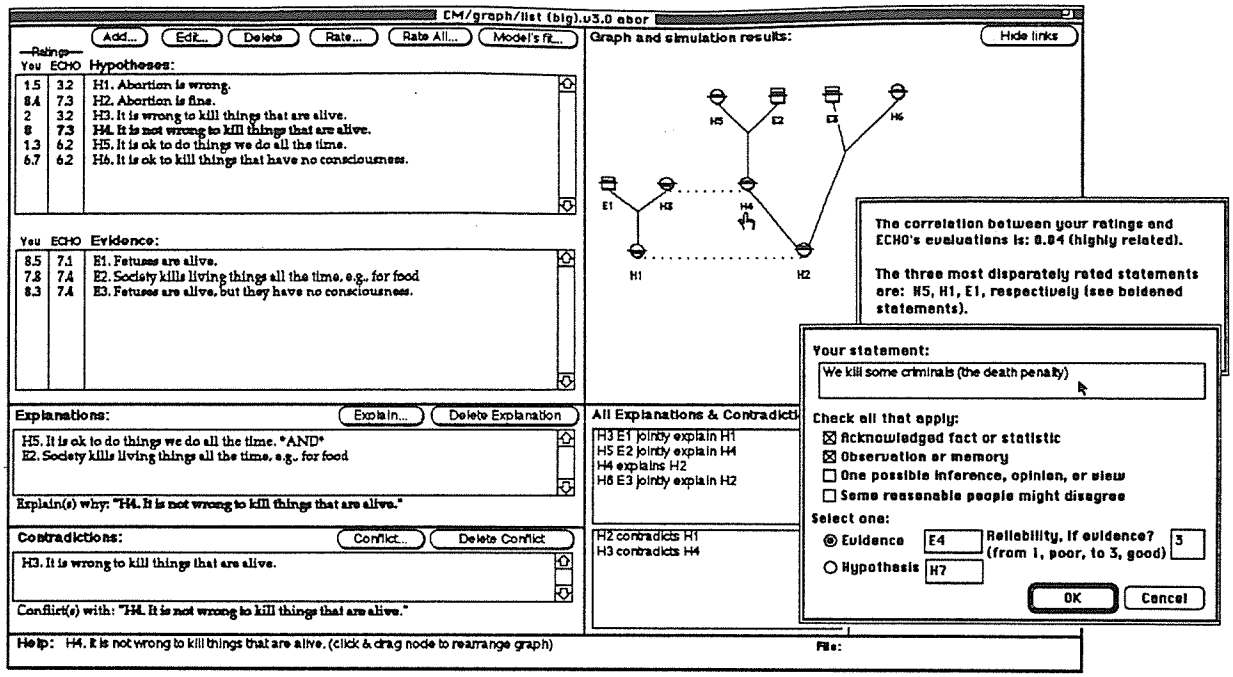


Figure 1. A person adds a belief about abortion (bottom right dialog) in response to *Convince Me*'s feedback.

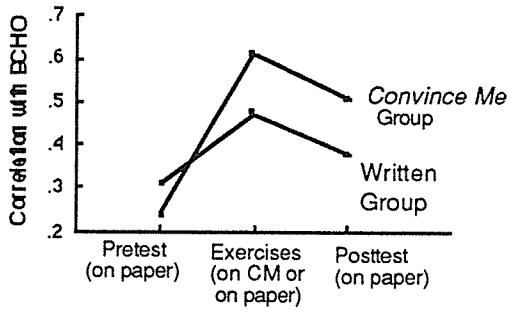


Figure 2. The ECHO model's overall fit, all arguments.

Subjects in both groups made about the same *total* changes to their arguments, but *Convince Me* users changed their argument structures twice as often as their ratings ($p < .05$). The trend was *reversed* for the Written Group, who changed their ratings twice as often as their arguments ($p < .05$). Users don't appear to view *Convince Me* as just a game they try to win by changing their ratings. On the contrary, *Convince Me* users seem more likely to reflect on and change the fundamental structures of their arguments.

DISCUSSION

Convince Me itself appears to help people coherently structure and revise their arguments, even beyond the enhancements offered by the curriculum. That is, the full system is both an effective "reasoner's workbench" *and* a learning environment that yields transfer to situations unsupported by the software and its attendant feedback. Future work will involve evaluating the utility of *Convince Me*'s argument listing and diagram representations, modeling human processing limitations [2] in *Convince Me*, and continued analyses of the nature of scientific reasoning.

ACKNOWEDGEMENTS

We thank Christine Diehl, Jonathan Neff, and the Reasoning Group for their help and suggestions.

REFERENCES

1. Cavalli-Sforza, V., Moore, J., & Suthers, D. Helping students articulate, support, and critisize scientific explanations. Proc. of the World Conference on AI in Ed., 1993.

2. Hoadley, C., Ranney, M., & Schank, P. WanderECHO: A connectionist simulation of limited coherence. Proc. of the 16th Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum, 1994.

3. Ranney, M. 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, in press.

4. Ranney, M., Schank, P., Hoadley, C., & Neff, J. "I know one when I see one": How (much) do hypotheses differ from evidence? Proc. of the 5th Annual American Society for Information Science Workshop on Classification Research, 1994.

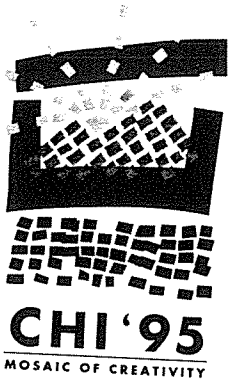
5. Schank, P., & Ranney, M. The psychological fidelity of ECHO: Modeling an experimental study of explanatory coherence. Proc. of the 13th Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum, 1991.

6. Schank, P., & Ranney, M. 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, Hillsdale, NJ: Erlbaum, 1992.

7. Thagard, P. Conceptual Revolutions. Princeton, NJ: Princeton University Press, 1992.

HUMAN FACTORS IN COMPUTING SYSTEMS

CHI '95 CONFERENCE COMPANION



CONFERENCE ON
HUMAN FACTORS IN
COMPUTING SYSTEMS
MAY 7-11 1995
DENVER COLORADO USA

Editors
Irvin Katz
Robert Mack
Linn Marks

Stacey Ashlund
Eric Bergman
Ruven Brooks
Larry Diamond
George Engelbeck
Wayne Gray
Jonathan Grudin
Muneo Kitajima
Angela Lucas
Catherine Marshall
Joy Mountford
Paul Reed
Tracy Roberts
Kurt Schmucker
Catherine Wolf