Ranney, M., & Schank, P. (1998). Toward an integration of the social and the scientific: Observing, modeling, and promoting the explanatory coherence of reasoning. In S. Read & L. Miller (Eds.), *Connectionist models of social reasoning and social behavior* (pp. 245-274). Mahwah, NJ: Lawrence Erlbaum.

Toward an Integration of the Social and the Scientific: Observing, Modeling, and Promoting the Explanatory Coherence of Reasoning

Michael Ranney
University of California, Berkeley

Patricia Schank
SRI International

It may seem odd for two cognitive scientists, each with little specific expertise in social psychology, to present a chapter that focuses on social cognition. Indeed, our past work may seem much more in the realm of scientific reasoning than in that of social reasoning. But one question that we have been asking, both of ourselves and of our colleagues, is, "What is the difference between 'scientific reasoning' and plain old 'reasoning'?" Generally, people hem and haw when confronted with this question, then speak of the latter as if it were social reasoning—and quite often, they mention socially based ruminations that involve suboptimal decisions, faulty heuristics, and inappropriately biased values, goals, and the like (see Gigerenzer, 1991; Tversky & Kahneman, 1974, and many others). Useful follow-up questions to such respondents include, "Well, is the difference between these two sorts of reasoning qualitative or quantitative?" Put another way (as many—including Einstein, 1950,—seem to have occasionally wondered), "Is scientific reasoning just (a) more likely to employ formal tools (like deduction or mathematics) and/or (b) more likely to involve the vigilant search for disconfirmation—something that just plain folks (Lave, 1988, p. 4) do, but less frequently?

Put rather bluntly, we have not been able to reject the hypothesis that the word "scientific" in "scientific reasoning" is superfluous. In an era of specialization, we realize that it is a bit out of fashion to undifferentiate reasoning (although one can argue that interdisciplinary cognitive science itself similarly bucks the trend); still, we are more struck by how much of the everyday is found in scientific reasoning (and vice versa) than by how unique scientific reasoning is. Thus, we believe that the principles of reasoning that have been seen primarily as characterizing scientific reasoning can equally well be viewed as central to social reasoning .[1] Bifurcating the set of reasoning processes into the social and the scientific is a bit like bifurcating a deity and still considering the encompassing religion to be monotheistic. In essence, adding either the modifier "social" or "scientific" seems unnecessary, unless one speaks about the *domain* being reasoned about (discussed later).

Consider the following experience. While at a social gathering, a friend's father remarked that sugar-laden foods dramatically boost children's activity. It was obvious, he claimed, because he had watched a birthday party full of kids "go crazy" after eating the cake. Another guest (perhaps a "better reasoner" on this topic) pointed out a glitch with the father's logic: Several factors other than sugar intake could account for the craziness (e.g., the party atmosphere, the greater disarray in the party room, relaxed supervision, etc.). The critic was essentially reasoning more completely and/or "more scientifically," bringing up issues of control or baseline conditions, alternative hypotheses, critical tests, and covariation detection (see Nisbett & Ross, 1980). (Note, though, that we do not suggest that the father was thus reasoning "more socially" because he failed to initially consider the critic's issues, as this might imply that the critic was *not* reasoning socially.)

---

[1] We thank Stephen Read for help in phrasing this premise. As he and Marcus-Newhall (1993) pointed out, "if these principles are as central in everyday social inference as they are in scientific inference, then it is critical that we gain a better understanding of their role in social inference" (p. 430; also see Fletcher, 1993).

Scientists often reason the same way the father did—with biases, untested assumptions, and observations that have no control situations to limit confounding (e.g., they use "fallback heuristics;" cf. Fletcher, 1993, p. 257). However, we believe that scientists appear scientific to the degree to which they spend a higher proportion of time engaging in sciencelike activities, such as criticizing hypotheses (both their own and others'), developing and calibrating precise measuring instruments and methods, engaging in formal (e.g., mathematical, computational, logical, and quasi-logical) analyses of a domain, soliciting extraneous reviews of their work, and the like. We further suggest that the more one has useful, auxiliary, physical, or social artifacts (e.g., calculators, computers, word processors, reasoning engines, logical systems, jargon), the more one can (at least appear to) "act scientifically" (see Fletcher, pp. 260, 265).

Naturally, no one engages in these behaviors 100% of the time, and there are several reasons for this gap between competence and performance. First, many scholars have noted (in one way or another) that we are limited processors of information, and such temporal, memorial, and computational demands strongly limit the rigor of our thoughts (e.g., Hoadley, Ranney, & Schank, 1994; Ranney, in press; Simon, 1955). To be globally coherent or rational means that one would ruminate so much as to never get out of bed in the mornings due to the exhaustive depth of processing required and the slow speed of our cognitive processors. (And even if one could rouse oneself, she would probably not be much fun.) Again, to the degree that our tools can reduce such processing limitations (for instance, an auditory alert before one's car runs out of fuel), our best-laid plans will indeed seem fairly scientifically (e.g., rigorously) laid out. Put more colloquially, scientists seem like eggheads mostly because of the "cool toys" they can use to study phenomena—in which toys can be both physical objects and special-purpose cognitive representations (Norman, 1988, 1993). Fletcher (1993) provided a similar perspective on processing limitations, grounded more in the literature of social and personality psychology, as to why, at different times, "people are both rational and rationalizers" (p. 255-260; cf. Ranney, 1996, on consistency and rationality).

As is often useful, it is worth reminding ourselves that scientists live in social worlds, just as social people live in scientific worlds (e.g., with at least some of the methods and products of science), regardless of whether either group likes it. Indeed, it is difficult to generate a reasoning scenario that does not involve, to some degree, both the social and the scientific in some way—that is, a purely social or purely scientific situation.[2] This is elaborated upon later when we discuss the relationships between, for instance, values (or goals), and hypotheses. For a quick example, though, note that the value (or at least what seems to be a value), "Don't act like an idiot" can be expressed as the hypothesis "Bad things happen when you act like an idiot" and vice versa.[3] Similarly, the hypothesis "The future will be better if we keep other organisms alive" might be seen as the value (e.g., the commandment) "Do not kill." The language of values often makes us think of social aspects, whereas the language of hypotheses often makes us think of scientific aspects. We suggest that these may be the same basic representations, but with differing perspectives, or "cover stories," if you will.

As relative outsiders to social psychology, we feel fairly unconstrained to point out different meanings of "social," and that some of these meanings differ from what is usually considered normative in the literature of social/personality psychology (e.g., as involving personality traits, group interactions, judgment, and decision making; cf. Fiske & Taylor, 1984). In this chapter, we highlight three readings of "social." As pointed out earlier, in one reading, some people (at least casually) contrast the social with the scientific, suggesting that social reasoning is a reasoning

---

[2] After all, even human languages, which often mediate reasoning processes, result from social consensus (Putnam, 1981), yet they have formal properties as well. In a similar vein, it would seem incorrect to argue that social reasoning does not involve the consideration of constructs that are generally seen as scientific, namely hypotheses or evidence.

[3] Such hypotheses may vary in concreteness or level of analysis. For instance, "People look down on you when you act like an idiot" might replace or supplement the more abstract "Bad things..." hypothesis.

"style" that has, in principle (i.e., not just practice), less empiricism, less objectivity, less rigor, less accountability,[4] or more emotion. We tend to reject this interpretation, as it seems to merely—that is, nonproductively—substitute "social reasoning" for "poor reasoning" and/or borrow from the following two readings.[5] (Fletcher, e.g., 1993, nicely addresses similar issues, and he occasionally links the "rational" with the "scientific" in his discussions.)

A second reading of "social" regards the *domain* of reasoning; in this form, social reasoning means reasoning about (small-scale) interpersonal situations (e.g., Miller & Read, 1991; Read & Miller, 1993). (This will be our default, but certainly not exclusive, meaning for "social.") Such reasoning might be about whether a certain girl will date a certain boy or how to act regarding an HIV-infected child (e.g., Ritter, 1991; see later discussion). A third meaning of "social" involves us as social creatures within a larger, communal, society—a society that can provide us with both great benefits and terrible hardships. The second and third meanings are occasionally seen as the material for commonsense or "folk" psychology. Fletcher contrasts such folk theories with ("scientific") psychological theories, in that they "have a wider range of uses and aims" and "consist of a more amorphous, flexible and sprawling set of concepts and models ..." (1993, p. 265).

---

[4] For example, some people view social reasoning as "cocktail conversation," in which even fairly outlandish and unsupportable statements may be generated when one is not held to account for them. (Tetlock, 1985, 1992, made similar points with respect to the fundamental attribution error and other phenomena.) This might be seen as a *mode* of reasoning to some people—an informal and off-the-cuff mode—but it is hardly the hallmark of good reasoning (as discussed later; cf. Putnam, 1981). (We thank Todd Shimoda and Bob Branstrom for pointing out this dimension to us.)

[5] This is why we use "socio-scientific" (at times) in this chapter. The reader may find it illuminating to try to generate a counterexample to this claim; that is, we invite the reader to try to find an instance of reasoning that is (a) poor *because* it is social and/or (b) does not conflate the following two notions of "social."

In this chapter, each of these three meanings of "social" is intermittently addressed, particularly with respect to our research regarding explanatory coherence, a system that was primarily developed to better study and analyze what we conceived of as "scientific" reasoning (Ranney & Thagard, 1988; Thagard, 1989). We begin with a summary of our past empirical work regarding explanatory coherence, in which we highlight aspects of reasoning about social situations and societal issues. This is followed by a more in-depth description of our studies involving a particularly controversial societal issue—that of the abortion debate—that we believe includes both social and scientific aspects. We then return to further analyze the purported distinction between social and scientific forms of reasoning, as well as distinctions among various component terms such as values, goals, hypotheses, and candidate actions (e.g., to consider whether the plurals "forms" and "terms" are needed). To reiterate for now, though, we believe that the basic principles and forms of social and scientific reasoning are essentially the same; what may differ between the two are aspects such as the amounts (and/or subtypes) of time, effort, systematicity, and cognitive tools used. Toward this chapter's end, we discuss a set of intended future directions for our research, which takes the shape of trying to answer the (perhaps ultimate social) question, "What should we be studying, such that society will maximally benefit?"

Studying and Modeling Explanatory Coherence

Much of our work has focused on assessing and applying the Theory of Explanatory Coherence (TEC) and its associated connectionist model, ECHO, which offer an account of how people evaluate the plausibility of beliefs comprising an explanation or argument (e.g., Ranney & Thagard, 1988; Schank, 1995; Thagard, 1989). TEC is comprised of roughly 10 principles that both establish local pairwise relations among cohering and incohering propositions and that empirically appear to play important roles in evaluations of the quality of an explanation (e.g., Read & Marcus-Newhall, 1993; Schank & Ranney, 1991). For instance, TEC assumes that the plausibility of a belief generally increases with (a) the simplicity with which it is explained (e.g., having fewer necessary explaining cohypotheses), (b) increasing breadth (i.e., offering greater coverage of observations), and (c) decreasing competition with alternative (especially entrenched)

beliefs. TEC is quasi-positivistic in that, all other things being equal, evidence tends to have "data priority"—that is, more plausibility—than do hypotheses. In addition, TEC has several holistic characteristics; for instance, (a) coherent beliefs are seen as *symmetrically* supportive, (b) incohering beliefs are seen as *symmetrically* conflicting, (c) coexplaining propositions support both an explanandum *and* each other, and (d) the overall coherence of a configuration of beliefs is "the best that it can be" (cf. the Gestalt Law of Prägnanz), given the pairwise (cohering or incohering) interactions of such beliefs (i.e., an argument reaches a state of minimal discord). TEC generally represents a somewhat uncontroversial view of explanatory coherence,[6] although some philosophers of science might quibble with a principle or two (e.g., the notion of data priority or its prevalence). Others offer somewhat different principles, albeit perhaps for somewhat different cognitive realms (e.g., on the epistemic criteria used for theory evaluation; Fletcher, 1993, pp. 254-255).

What considerably distinguishes TEC from other such systems is its embodiment in a computational, constraint-based, model. In essence, such models are homeostasis-seeking devices, driven to find points of equilibrium or minimal energy—in this case, among beliefs in a system of beliefs that may be quite "messy" (e.g., involving multiple contradictory/competitive elements that are centrally embedded in an argument). In this way, constraint-satisfaction models reify what have been, prior to computational modeling, more "hand-waving" notions such as balance theory, impression formation, Gestaltism, and cognitive consistency or dissonance (see Gabrys, 1989; Kunda & Thagard, 1996; Read & Marcus-Newhall, 1993; Read, Vanman, & Miller, in press; Shultz & Lepper, 1992, 1996; Spellman, Ullman, & Holyoak, 1993; Wertheimer, 1982). These were generally promising models of various high-level (and other) forms of thought that were limited by the lack of formal (e.g., computational) ways in which to assess them.

---

[6] TEC is more controversial as an account of scientific inference, in which the main competitor is the Bayesian approach (see Thagard, in press, for a comparison of ECHO with probabilistic models of inference).

In contrast, connectionist implementations of such constraint-satisfaction systems are computational and considerably more formal. They can generally be thought of as a comingling of constructs or beliefs (nodes in a network) that are, using an interpersonal metaphor, either "buddies" or "enemies" of each other (but usually not both). The "friendship" of the buddies takes the form of excitatory constraints/links, such that friends generally either do well together or poorly together. On the other hand, the "distaste" of the enemies represents inhibitory constraints/links that tend to drive each other's acceptability (or "activation") down. As activation ("popularity") settles among the various beliefs, this computational currency of evaluation will indicate the relative plausibility of each of the beliefs (e.g., which cliques and subcliques of propositions are most exalted). (Also see Bereiter, 1991, and Thagard, 1996, for other descriptions of such implementations.)

TEC's model, ECHO, is based on the claim that beliefs are related explanatory entities, and evaluating their plausibility is an interactive, principled, coherence-seeking process (Ranney, in press; Thagard, 1989). Belief evaluation in ECHO involves the satisfaction of many constraints, determined by the explanatory relations among propositions and a few processing parameters. ECHO employs a connectionist architecture in which each node represents a proposition (hypothesis or piece of evidence) with an associated activation value, and each connection represents an explanatory relation (explanation, contradiction, or competition[7]) with an associated weighting.[8] ECHO's connection weights and propositional relationships are provided, depending upon the methodology employed, by default, by the experimenter, or by the subject (e.g., Ranney & Schank, 1995). For example, consider the following scenario, which involves two different viewpoints about abortion (from Schank, 1995):

---

[7] "Competition," as used here, is a technical term (Thagard, 1992), and is implemented in a variant of ECHO, called ECHO2. However, its utility is questionable (see Ranney, Schank, Mosmann, & Montoya, 1993), so it is not used or discussed here.

[8] More complete descriptions of ECHO's algorithms are available elsewhere (e.g., Thagard, 1989, 1992).

> Smith believes that abortion is wrong because fetuses are alive. Jones disagrees, saying that abortion is fine, because we as a society kill living things (e.g., for food) all the time. (p. 214)

Given the following possible encoding (e.g., by a subject) of this argument,

```
hypothesis H1 "Abortion is wrong"
hypothesis H2 "Abortion is fine"
hypothesis H3 "It's wrong to kill things that are alive"
evidence E1 "Fetuses are alive"
evidence E2 "Society kills living things all the time"
H1 contradicts H2
E1 and H3 jointly explain H1
E2 explains H2
```

ECHO would accept this input and generate the network shown in Fig. 8.1, in which the solid line between E2 and H2 represents an independent explanation, the solid Y-shape represents a joint explanation, and the dashed line represents a contradiction.

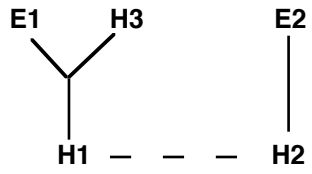

FIG. 8.1. A simplified representation of a small ECHO network.

Node activations represent propositional acceptabilities, and are initially assigned to zero (neutral). (Final activations range from -1, complete rejection, to 1, complete acceptance.) Links in ECHO are symmetric, and the link weights (which also range from -1 to 1) are specified by three processing parameters—*excitation*, *inhibition*, and *data excitation*. Excitation determines the weights on links between cohering propositions (with the value divided by the number of explanatory propositions in the case of a joint explanation), inhibition determines the weight on links between incohering propositions, and data excitation specifies the weight on links between data (usually evidence) and

the "special evidence unit" (SEU, a unit with activation set at a constant 1.0, to give data a bias toward acceptability). The data priority of a particular piece of evidence may also be specified separately, if desired. Continuing with our example above, let us assume that the parameter settings are excitation = .03, inhibition = .06, and data excitation = .055 (our standard default settings). Given these values, the ECHO network shown in Fig 8.1 would be more completely represented as shown in Fig. 8.2, in which the numbers in italics represent link weights, and the numbers in parentheses represent initial node activations.

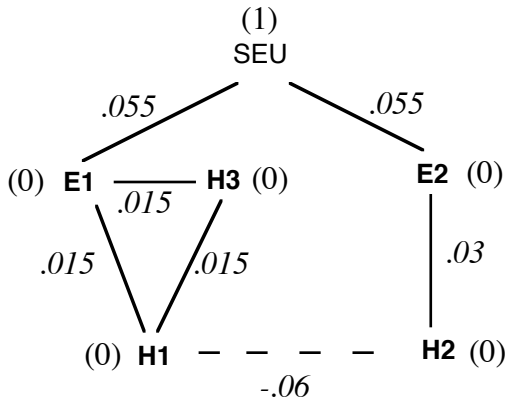

FIG. 8.2. A more explicit initial representation of the ECHO network in Fig. 8.1, with connection weights and initial node activations.

After setting up the network, ECHO's constraint-satisfaction engine determines the acceptability of each proposition. 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 partially determined by a fourth critical parameter—the *decay rate*—and the net input to the unit. Decay specifies the percentage of the activation that a proposition loses at each cycle (when otherwise unchanged). The net input to a unit u_j is the weighted sum of the activation of each neighboring unit u_i, in which a weight is the common link value w_{ij}. Each unit u_j is updated using the following equation (cf. Rumelhart & McClelland, 1986):

$$u_j(t+1) = u_j(t)\,(1-decay) + \{\ net_j\,(max - u_j(t) \text{ if } net_j > 0 \quad \text{and}$$

$$net_j\,(u_j(t) - min) \text{ otherwise}\}$$

in which $net_j = \sum_i w_{ij}\ u_i(t)$

Although ECHO has a few other processing parameters, the four presented here (excitation, inhibition, data excitation, and decay) appear most critical to ECHO's performance (Schank & Ranney, 1991). Returning to our example, the network (in Fig. 8.2) takes 149 cycles to settle, with final activations shown in Fig. 8.3. Note that, for this representation, ECHO essentially predicts that the modeled subject will find that proposition H2 is moderately acceptable (.47 activation) whereas H1 is mildly rejectable (-.29). (For simplicity, link weights are no longer shown because they do not change; the SEU does not change, either, because it is "clamped" at 1.0).

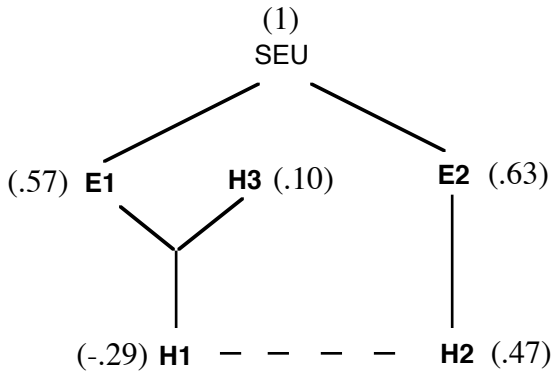

FIG. 8.3. Final activations values (in parentheses) for the network shown in prior figures.

Thus, TEC offers a model (ECHO) that helps us understand how people draw inferences in social environments—one that yields testable and reasonably precise predictions. For example, people might be said to accept hypotheses about others on the basis that these hypotheses are coherent with their current beliefs, and yield coherent explanations of observed behavior (Thagard, 1989). Pennington and Hastie (1988), as well as Thagard, claim that explanatory coherence plays a

crucial role in jurors' decision making, and Read, Miller, and Marcus-Newhall (Miller & Read, 1991; Read & Marcus-Newhall, 1993; Read & Miller, 1993) have used the principles of explanatory coherence to understand and guide models of social interaction and relationships.

Early assessments of ECHO's modeling effectiveness focused on the post hoc modeling of arguments extracted from texts (e.g., scientific treatises and juror reasoning, Thagard, 1992; and social interactions, Miller & Read, 1991) and on subjects' verbal protocols (e.g., regarding their conceptions of physical motion; Ranney & Thagard, 1988). More recently, we have used ECHO to (a priori) predict the strength of subjects' beliefs (see Ranney & Schank, 1995). Initially, we used ECHO to predict subjects' text-based believability ratings (Ritter, 1991; Schank & Ranney, 1991). This methodology involved subjects reading some paragraphs that embodied a particular ECHO network topology; subjects were then asked to rate, on a believability scale, each proposition that was embedded in a text. Texts involving social relationships were particularly salient in Christopher Ritter's work in our laboratory (e.g., Ranney, Schank, & Ritter, 1992). In the following example, between-gender social relations were highlighted:

> A boy wants to ask a girl to see a movie with him. Will Emily say yes or no to Zachary?
>
> On the one hand, Zachary believes that Emily may dislike him. Emily laughed at him when he fell on the baseball field. Emily did not talk to Zachary when he saw her at the mall. And when Zachary ran for class vice-president, Emily supported Zachary's opponent. The possibility that Emily dislikes Zachary would mean that she will say no to seeing a movie with him.
>
> On the other hand, Zachary believes that Emily might indeed like him. Emily attends his baseball practices frequently. Sometimes Zachary catches her watching him in class. And he got a valentine from Emily in February. Finally, the assumption that girls are more prone to like boys than to dislike them suggests that she might like Zachary. The possibility that
> Emily likes Zachary means that she will say yes to seeing a movie with him.

What do you think? (Ritter, 1991, p. 47)

In initial trials with this scenario, we ended the text with “Should Zachary ask Emily out?” instead of “What do you think?,” but biases from subjects' extraneous knowledge indicated that he should ask her out if there were *any* reasonable chance (even significantly less than 50%) that she would agree to it. Thus, the result of comparing ECHO's modeling of this scenario with the believability ratings subjects offered for the text's embedded propositions, and in combination with subjects' verbal protocols, indicated the following: Subjects were clearly bringing extraneous background knowledge and (perhaps past dating) experiences to the situation, reasoning akin to jurors concluding that the “defendant” is “guilty” (cf. Emily saying "no" to a date) only if there were no “reasonable doubt.” Schank and Ranney (1991) found similar results, with subjects bringing considerable (and often social) background knowledge to bear, sometimes resulting in divergent interpretations of what we incorrectly assumed would be fairly straightforward linguistic entities such as "think" and "believable."

Such text-based assessments of ECHO were largely successful, modeling even dynamically changing belief evaluations while yielding model-fit correlations, between ECHO's activation predictions and subjects' observed belief evaluations, of up to 0.8 (Schank & Ranney, 1991). This thread of research also allowed for the assessment of TEC's principles (also see Ranney, Schank, Mosmann, & Montoya, 1993; Read & Marcus-Newhall, 1993). Even so, subjects were clearly considering extratextual background knowledge in their deliberations, which was (again) often social in nature. Consider another, larger (full-page), scenario, in which subjects were asked to reason about whether it was safe to send children to a preschool in which another child was tested positive for HIV (Ritter, 1991). Some of these subjects mentioned that preschool children that they knew often engaged in noncasual behaviors (e.g., biting), which dramatically changed their representations of the situation. One said (pp. 22-23):

> I had some strong feelings on this one. Kids bite each other, in preschool. Often. A lot. It’s not casual. Both my kids were bit [sic] in their preschool education. Both

> my kids—they're good kids, well adjusted—bit other kids. Most do, at some point, bite another kid.

Similarly, many subjects dwelled on how the HIV-infected child or his or her parents would feel about what could essentially turn into a discriminating quarantine, adding concerns such as (p. 23) "I'm very worried about our society isolating HIV-infected people. I'm very concerned about that."

In cases such as these, subjects showed considerably lower correlations between ECHO's activations and their believability ratings. In essence, the set of textual propositions and the text itself was not fully representative of the information that the subjects were considering (due to their extraneous background knowledge), so the fits of the models suffered.

Responses such as these "biting" and "discrimination" ones inspired us to develop a novel method ("bifurcation/bootstrapping"; Schank & Ranney, 1992) to elicit and account for such background knowledge through an on-line interview/protocol session. We concurrently assessed the intercoder reliability of the representations encoders generated as input to ECHO by having multiple encoders "translate" subjects' verbal protocols into (evidential or hypothetical) propositions and (explanatory or inhibitory) relations among the propositions. In this case, we investigated ECHO's ability to model individual subjects' beliefs about physical motion. Our protocol modeling results indicated both reasonably good data fitting and intercoder reliability. Belief revision over time was also well modeled, as were attentional and memorial constraints (e.g., Hoadley et al., 1994; Ranney et al., 1993).

Most recently, our desires to automate the explication of individuals' knowledge bases and belief assessments, and to aid students in articulating and revising their theories, led us to develop our "reasoner's workbench," as described in the next section (Schank, 1995; Schank & Ranney, 1993). This system, Convince Me, is meant to both study and foster aspects of reasoning that are commonly attributed to scientists, for instance: (a) a more articulated and conscious awareness of hypotheses, evidence, and inference, (b) a more systematic way to think about arguments, and (c)

more explicit links among beliefs. However, because we claim that these reasoning aspects are hardly unique to scientists, we address them with the aid of some social situations in the following section.

Promoting (While Observing) Explanatory Coherence

The *Convince Me* System and Some Social Applications

Convince Me is a domain-independent computational "reasoner's workbench" program that supports argument development and revision, and provides ECHO-based feedback on the coherence of subjects' articulated beliefs. (The system has been employed in diverse realms, e.g., biology, physics, geography, linguistics, etc.; Ranney et al., 1993; Schank, 1995; Schank & Ranney, 1991, 1992; Schank, Ranney, & Hoadley, 1995). The associated curriculum discusses distinctions between hypotheses and evidence, strategies for generating and evaluating arguments and counterarguments, and reasoning biases and how to reduce them. Using Convince Me, individuals can (a) articulate their beliefs regarding a controversy, (b) categorize each notion as being either evidential or hypothetical, (c) connect their beliefs explanatorily and/or inhibitorily, (d) provide ratings to indicate the believability of each statement, and (e) run the ECHO simulation to obtain various forms of feedback. Convince Me also incorporates other important features, such as a tool for diagramming an argument's structure, as well as support for modifying one's arguments, belief ratings, and even the parameters that govern ECHO's "reasoning engine."

Figure 8.4 shows a subject's argument on the aforementioned abortion controversy (one of many in the curriculum) in Convince Me. After entering an argument, specifying believability ratings, and running an ECHO simulation, the networked argument's graphed nodes reflect ECHO's activations in a "thermometerlike" fashion (see Fig 8.4, upper right). The higher the mercury, the more ECHO accepts the statement; the lower the mercury, the more ECHO rejects the statement. (Likert-scaled numerical equivalents are also provided; see the upper left of Fig 8.4.) In addition, Convince Me can report a "model's fit" correlation between one's believability ratings and ECHO's scaled activation values, indicate how related the two sets of ratings are (e.g., "mildly opposed,"

"moderately related," "highly related"), and highlight the (three) pairs of values that differ the most (see the middle box in Fig 8.4).

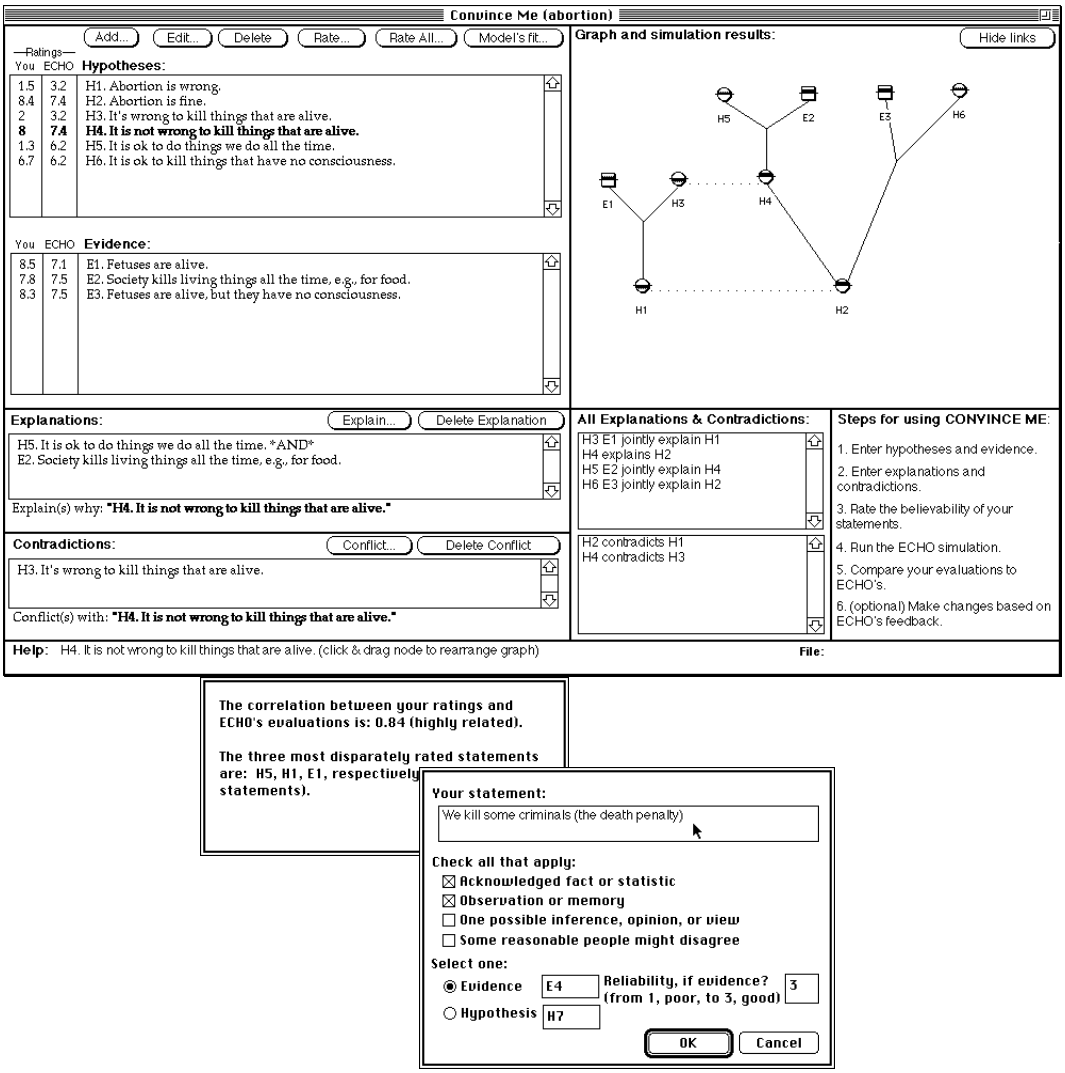

FIG 8.4. A subject adds a belief related to the abortion controversy ["We kill some criminals (the death penalty)"] and classifies it as reliable evidence (bottom) in response to Convince Me's feedback (middle). Proposition H4, a central node, is highlighted in the main screen area.

Convince Me has been used by subjects to reason about various other social situations, such as interpretations of human behavior (e.g., whether yawning indicates a subconscious display of aggression or just a lack of oxygen), and whether the use of marijuana should be legalized. The following are two sample texts on these topics, presented to subjects who were using the system (Diehl, 1995; Diehl, Ranney, & Schank, 1995):

(a) "Kathryn thinks that marijuana should be legalized because if it were, then it could be regulated, and drug crimes would decrease. Phyllis thinks that marijuana should not be legalized because then it would be used more and lead to further degradation of our society."

(b) "Wanda and Dave are walking through Pinetown one night, and both notice that an approaching teenager yawns when passing them.

Dave thinks that the teenager's yawn was an subconscious aggressive display. He learned in biology that humans are genetically close to apes, and ape studies suggest that apes engage in "threat yawns." In a group, dominant male apes yawn more—an action that shows off their long canine teeth—while subordinate apes more often cover their yawning mouths with their paws. He says that since Pinetown is a dangerous area, this would explain why the teenager yawned when passing them.

Wanda disagrees with Dave. She notes that people, as well as non-primates such as dogs, yawn when they are alone as well as in groups. She has read that yawning provides more oxygen to the brain and that the more oxygen, the more glucose we can burn for energy. She thinks that since it is late, the teenager is probably tired and yawned to get more oxygen to stay alert. She claims that the hypothesis that yawning is to increase oxygen also explains why it *seems* contagious—people in the same room are all just breathing the same stuffy air, and all need more oxygen."

As we expected from our prior research, subjects generated various alternative hypotheses and extraneous evidence (i.e., that went beyond the given text) regarding these situations. In the marijuana scenario, for example, one subject came up with (and modeled) additional beliefs such as "Marijuana should be legalized for the medical field and not for recreational use," "Patients would benefit from marijuana treatment," and "The economy would benefit from the regulation of marijuana."

Not surprisingly, subjects generated slightly fewer additional beliefs when reasoning about the more elaborated (and perhaps less value-laden) "yawning" scenario. These were usually constrained to why the teenager might be tired (e.g., "The teenager just finished a paper for class"), other reasons for yawning (e.g., boredom, or to alleviate congestion), and potential biases of the viewers ("Wanda is a biologist," "Dave is an anthropologist"). On average, subjects' representations for this argument included four hypotheses, eight pieces of evidence, 18 explanations (four of them joint explanations), and seven contradictions—indicating that the arguments they generated were quite complex. Average belief-activation correlations for subjects using Convince Me to complete this exercise were fairly high at .70, and correlations ranged from .73 to .88 for the other exercises completed using the system (Schank, 1995). These results are in concert with our earlier findings that ECHO can effectively predict subjects' believability ratings for both small- and medium-sized text-based arguments (e.g., Ritter, 1991; Schank & Ranney, 1991), as well as the generally larger subject-generated arguments gathered via our bifurcation/ bootstrapping verbal protocol method (Ranney & Schank, 1995; Schank & Ranney, 1992). Completing exercises with the system using scenarios such as this one also appears to improve subjects' abilities to discriminate between the notions of evidence and hypothesis, as discussed more later (and in Ranney, Schank, Hoadley, & Neff, 1996).

Abortion: An Elaborated Example Employing the *Convince Me* Curriculum and System

We recently assessed (e.g., Schank, 1995) Convince Me's effectiveness as a system for fostering reasoning skills by contrasting subjects' performance under two conditions with otherwise identical written tests, instructions, and curricula: instruction involving (a) Convince Me versus (b) no software—a completely on-paper curriculum (the "Written Group"). The abortion example (shown earlier) was one of the four textual stimuli used. Others included competing theories regarding yawning (described earlier), medical diagnosis, and various predictions about pendular release trajectories (from Schank & Ranney, 1992). Results indicate that Convince Me users' beliefs were significantly more in accord with the structures of their arguments (even prior to feedback from the ECHO model), as evidenced by higher belief-activation correlations, compared to the Written Group (e.g., .81 vs. .34 for the abortion text).[9] Further, Convince Me users changed their argument structures twice as often as their ratings, whereas this effect was *reversed* for the Written Group, who changed their ratings twice as often as their arguments. Thus, Convince Me users apparently do not just try to "mimic" ECHO by changing their ratings; on the contrary, compared to students developing arguments on paper, students using the system seem more likely to reflect on and change the fundamental structure of their arguments. Convince Me subjects also tended to employ more explanations and contradictions in their arguments. These results generally explain why Convince Me users' belief-activations correlations were higher—namely, that more explicated and revised arguments should better reflect underlying beliefs.

In addition to assessing the prescriptive utility of Convince Me, we are interested in new descriptive questions that bear upon social psychology, such as, "How do individuals' values affect the coherence of their argument?" For instance, are our abortion findings at all modulated by the side one takes in an argument (e.g., prolife vs. prochoice)? Also, when subjects generate their own propositions, are they proficient at categorizing them as hypothesis or evidence—or are their categorizations questionable? To address such questions, a more detailed analysis of the abortion argument was conducted, because it was the last of four final exercises subjects completed (i.e., they

---

[9] This was also the case across all four textual stimuli.

were "warmed up" with the system), and the textual stimulus provided was minimal (two sentences, with about four or fewer propositions; see the earlier text above) so that the majority of the argument came from the subject rather than the text.

Analyses of the abortion data revealed that prolife and prochoice leanings, as evidenced by their central believability ratings, were not correlated with (Convince Me or Written) group membership, nor were belief-activation correlations related to the side taken (e.g., prolife, prochoice, or neutral) in the argument. (Of 20 subjects, 10 were prochoice, 5 prolife, and 5 neutral.) Overall, subjects generated numerically (but nonsignificantly) more propositions for their side of the argument (the ratio of "my side" vs. "other side" propositions was 9:8; see, e.g., Perkins, 1995, on sides), and there were no significant differences between groups on this measure.

The questionability of subjects' categorizations also did not differ significantly between groups. In a blind categorization,[10] one of us questioned (in essence) the categorization of about 16% of the subjects' statements—about 8% of their hypothetical classifications, and about 23% of their evidential ones. However, the disagreements were not distributed evenly across subjects—20% (3 prochoice, 1 prolife) accounted for over half of the questionable categorizations, and 35% (4 prochoice, 2 prolife, 1 neutral) appeared to have no questionable categorizations at all. The most common difficulty occurred when a subject categorized an assertion as a piece of evidence when the researcher viewed it as a hypothesis. For instance, one subject classified the statements "Population is too high" and "Unwanted children only cause more problems" as evidence when both seem clearly arguable (partly because they include vague quantifiers). Another subject listed as evidence "It is a personal not societal issue" and "Everyone has a God-given right to live," which, again, are some of the more controversial (and arguable) assumptions at the center of the debate![11]

---

[10] Additional subject-generated data (e.g., from responses to the check-boxes shown toward the bottom of Fig. 8.4) support the researcher's blind categorizations (see Schank, 1995).

In sum, we did not find evidence to suggest that individuals' values (as shown by prolife or prochoice stances) affect the coherence of their argument. On the contrary, our subjects (somewhat surprisingly) did not offer significantly more statements for their own side of the argument, nor did the coherence between believability ratings and argument structures seem related to the side they chose. Some subjects seemed to take liberties in what they called (and failed to call) evidence, but these liberties did not appear to correlate with their (prolife or prochoice) viewpoint. One could credit our curriculum with reducing the biasing influence of such values in the way it fosters the articulation, interrelation, and revision of arguments; either way, the Convince Me software enhances the correlation between subjects' belief evaluations and the structures of their arguments.

Of course, other considerations should be entertained. For example, analyses of debates on value-laden issues other than abortion (e.g., situations that bring out other values, or have fewer "stock arguments," etc.) may yield different results. The abortion debate is often based on (possibly deeper) politico–religious values and/or the differential valuing of societal versus individual rights, and arguments for both sides are often familiar to subjects. Hence, it may be easier for them to simply reproduce common statements supporting each side of this debate. Similarly, the degrees of dogmatism associated with various values may play a role in the assessments of arguments. It may be that values that are based strongly on doctrine (as could be suggested by extreme ratings), when "unwrapped" into an argument, could lead to less coherence—in the form of beliefs that are either not well explicated, socially difficult to explicate, or are asserted but not well-supported by the offered data. As an example, suppose that part of a subject's support for the prolife stance involves an (even warranted) appeal to authority, such as "my father is prolife." In turn, support for this proposition might include "my father is usually right" and/or "my father was right about my no-

---

[11] Less common were questionable categorizations of hypotheses that seemed more like evidence, such as "Fetuses are alive" (they *are* made of living tissue—the issue is whether they are alive as people in some independent sense that warrants rights) and "We as a society kill living things" (we clearly kill plants and/or animals for food, we kill some criminals, etc.).

good ex." Subjects might feel inhibited to type in such propositions, though, and their arguments would thus be less explicated and less supported. However, note that we did not find much evidence for such dogmatism; only 14% of our subjects' ratings were at the extremes of the scale (i.e., "completely believe" or "completely reject") for the "Abortion is wrong" and "Abortion is okay" statements and all of these extreme ratings came from only 20% of the subjects.[12]

A Modest Proposition

Several recent applications of Convince Me concern our desire to promote both globally coherent reasoning (cf. Ranney, in press) and more environmentally conscious reasoning, a topic that we discuss a bit more toward the end of the chapter. As a societally relevant foreshadowing example, though, consider the proposition, "We should never cause the extinction of another species." Far from an ivory tower controversy, the debate around this hypothesis (or value) can easily reach from our Pacific Northwest, where the survival of the spotted owl is pitted against various economic (e.g., job security) interests, to Washington, DC, where logging industry lobbyists prowl Congress and the administration, to nasty pathogens located in the bowels of the Center for Disease Control in Atlanta.

One of us has particularly struggled with this controversy, and employed Convince Me to articulate his reasoning. The argument (which is currently too large to usefully portray here in its totality) reflects his confusion: Naturally, all other things being equal, it is better to have more species than fewer species. However, some species are quite deadly to humans. Should they not be destroyed? Certainly, our ancient ancestors would probably not be terribly sad to hear that the last

---

[12] Many other researchers have found that subjects sometimes avoid extreme rating values; this is the well-known "restricted range of variables" problem (e.g., Jensen, 1980). For instance, data collected in conjunction with Florian Kaiser and others indicate that our undergraduate participants from the University of California, Berkeley, behave more homogeneously, ecologically speaking, than a more heterogeneous sample of Swiss auto drivers (cf. Kaiser, Ranney, Hartig, & Bowler, 1997).

saber-tooth tiger passed on; will we be upset when the last smallpox microbe bites the dust—or when perhaps HIV and the "killer bees" are effectively extinct? The following (paraphrased) dialogue between one of us (M) and his 7-year-old daughter (R) highlights other questions regarding how much human comfort (as in the Pacific Northwest), or even individual lives, mean, in contrast with other species' survival. It was prompted by a James Taylor song ("Mona," 1985), in which the narrator practiced what we might call early euthanasia on his pig because she "got too big to keep and too damn old to eat."

R: Would you shoot me with a 12-gauge if we couldn't both live?

M: No, I'd probably kill myself first to let you have more food; you eat less, and I've already lived about half my life, so it wouldn't be fair for me to live instead of you.

R: Would you kill a pig instead of yourself?

M: Probably, but it depends upon how many pigs remain on Earth. And pigs are awfully smart and useful, compared to many animals.

R: Would you kill me instead of the last pig on Earth if you had to choose?

M: I think I hear your mom coming home! [This statement is highly paraphrased.]

Even intriguing exchanges such as these, which involve the contrast between personal, societal, and global interests, may be analyzed with the assistance of Convince Me. It is our hope that the system will promote both more extensive and more principled reasoning about social issues that we should all be considering. Again, this point is elaborated toward the end of this chapter. For now, we leave this "societal" reading of "social" and return to our criticism of the first reading, in which the social is contrasted with the scientific.

Equating the Social and the Scientific: Not Just Theoretical Imperialism

From the earlier discussion on the proposed unity of reasoning found in both the social and the scientific realms, we might be seen as domain chauvinists. After all, "is not leather the only thing to the cobbler?" A critical reader might suggest that we want to unify the social and the

scientific—in terms of hypotheses and evidence and with respect to excitatory and inhibitory relations—because we want our perspective to seem preeminent. As the kids say, "Well, duh!" Of course, we would want that to happen; who wouldn't? Indeed, to a considerable degree, this criticism-and-reply proves our point, as follows: We (as scientists) can of course think of our goal as "Unify apparently disparate phenomena." It follows from our prior, similar, argument regarding values, that having this goal is essentially the equivalent of entertaining the hypothesis, "Unifying disparate phenomena has desirable effects." Naturally, we truly believe that the unification we propose both works and is correct. Still, those who suspect that we are more Machiavellian might suggest that the aforementioned goal stands in for a more directly self-serving goal, such as "Gain more fame" or "Acquire more wealth." But each of these can similarly be converted into hypotheses by a small set of simple transformations (e.g., production rules; Newell & Simon, 1972), such as: "If goal x, then hypothesis 'it is best to x,'" and perhaps vice versa (cf. Putnam, 1981).

Once goals are seen essentially as hypotheses, configurations of other hypotheses and evidence can link with them in explanatory and contradictory/competitive ways, and what initially may have seemed like a value-laden decision is now seen as ("merely") a scientific knowledge base, involving observations and data, predictions and presumed consequences, theories and models (whatever those are)—or even facts.[13] Predictions that have the highest activations may be thought of as reflecting the actions that represent the best decisions currently being considered by the system.

---

[13] We are, to some degree, implicitly taking the philosophical stance of ethical naturalism here—for instance, that "is" statements can be brought to bear on "ought" statements in one's argument. We also understand that this is part of a rich and ongoing discussion in philosophy and linguistics regarding the purported (and increasingly unfashionable) fact-value dichotomy (e.g., Lakoff, 1987; Millgram, 1995; Putnam, 1981).

Note the parenthetical "whatever those are" above. This was not meant facetiously. In fact, very few of the terms that we have employed to discuss socio–scientific (or "value-theoretic") reasoning and decision making are even close to being precise ones. We empirically demonstrated, for instance (Ranney et al., 1996; Schank, 1995), that even the distinction between hypothesis and evidence is a fuzzy one. For both novices and experts in scientific reasoning, relatively low interrater reliabilities are obtained when subjects are asked to rate propositions with respect to their hypothesis likeness or evidence likeness. More recent data (Diehl, Castro, & Ranney, 1997) indicate that, although subject-generated exemplars of evidence and hypotheses have certain prototypical linguistic characteristics, what "fact" means—with respect to these other two constructs—is highly variable. For instance, is a fact a true piece of evidence, a highly supported hypothesis, or something else? We have obtained a multiplicity of such responses.[14]

Lest we dwell only on the fuzziness of the terms involved in scientific reasoning, we should point out that more socially oriented terms are probably no more clear. Earlier, we proposed that we can pretty much treat hypotheses, goals, values, and candidate actions as a single kind. In other words, there is little agreement about what distinguishes among these terms, and the "noise" swamps the quite variable "signals" of those who believe that they can truly discriminate among such terms (e.g., the variety found in Pervin, 1989; see especially pp. 473-474; also see Gollwitzer & Bargh, 1996). For instance, we propose that a single propositional node in an ECHOlike connectionist network (e.g., "We should reduce the population") might be seen as a goal to some theorists, a hypothesis to others, a value to a third group, and a candidate action to still others. Some theorists might base their categorization on contextual features or even some processing subtleties (e.g., regarding the proposition's embeddedness/connectivity or "priority bias"; cf. Millgram & Thagard, 1996), but there clearly exists a nonuniformity among these views.

---

[14] Although almost all of our subjects saw fact and evidence as different constructs, they often disagreed about whether one was a type of the other, or if so, which was the superordinate in the categorical hierarchy.

As an example of such diverse views, consider ECHO's sibling program, DECO, a similar constraint-satisfying model that is applied to multiaction planning (e.g., Millgram & Thagard, 1996). The theory underlying DECO makes a distinction between goals and candidate actions that seems highly fuzzy (an even fuzzier one than that between hypotheses and evidence, we believe) even though the distinction entails strong computational consequences in that goals gain considerable priority. For instance, in our modeling with ECHO, we have felt rather comfortable representing (albeit somewhat implicitly) goals and actions generally as hypotheses[15] [e.g., "I hypothesize that I should (take the action to) go to the library," or "I hypothesize that I should (have the goal to) meet Florence"; cf. Millgram & Thagard, 1996]. It has also seemed to us that ECHO, as DECO is intended to do, models decision making of sorts (e.g., whether one should decide that a defendant is guilty or innocent, or whether one should decide that some symptoms represent one disease rather than another; Ranney et al., 1993; Schank & Ranney, 1991; Thagard, 1989, 1992). Indeed, given that DECO and ECHO are so similar (e.g., near isomorphic in their principles, in terms of processing, semantics, etc.), it is not yet clear to us that a separate theory-and-model will, ultimately, be needed. (We have similar concerns regarding IMP, Kunda & Thagard's 1996 model of impression formation.)

Perhaps an even more social term, "values," provides still less terminological comfort (e.g., vs. "facts;" Millgram, 1995; Putnam, 1981). What is a "value," anyway? We propose that a value may be seen as a hypothesis that is both so supported by strong propositions (e.g., highly activated evidence and fellow hypotheses) and so distal from all but the weakest competitors, that it seems

---

[15] In essence, we do not see the need for a processing priority for goals. Instead, we would view high-priority goals as hypotheses that are very well supported by a complex of reliable evidence and very strong hypotheses. The goal/value, "It is best to stay alive" might be one such hypothesis (e.g., supported by the fear of death evidenced in our culture, the conservatism of avoiding change, the assumption that one will ultimately get to experience death anyway, etc.); even so, it might become rejected (e.g., in order to save the lives of loved ones).

virtually indubitable, almost like a special evidence unit (Thagard, 1989). Note that, as with goals, we have once again described a fairly socially oriented construct with our "scientific/theoretical fixings." Some will undoubtedly chafe at this, perhaps suggesting that value-laden issues cannot be the subject of scientificlike debate. However, it is our experience, such as with the abortion-debate analysis above (among others), that what appear to be value-laden issues can be captured to a remarkable degree in terms of evidence, hypotheses, explanations, and inhibitory relations. In other words, we remain to be convinced that values, like goals, are not already fundamentally captured by explanatory coherence theory.

In contrast to our view of values as propositional (also see Millgram & Thagard's "intrinsically valuable goals"), some researchers appear to see a value as subpropositional, that is, more like an abstract word (cf. Putnam, 1981) or a phrase rather than as a statement, linguistically. McCarty and Shrum's (1994) work provides an example. Although they acknowledged that the distinction between value orientations and personal values is slippery, they represented value orientations by longish phrases such as "Being a cooperative participant in group activities," and represented (from Kahle, 1983) personal values (and even evidencelike "behaviors") with shorter, one-to-four-word phrases such as "Security" or "Fun and enjoyment." In this paradigm, full propositions (e.g., sentencelike statements) take the form of attitudes, opinions, ideas, goals, and/or beliefs (although distinctions among these also seem slippery to us; cf. Gollwitzer & Bargh, 1996; Pervin, 1989; Scott & Willits, 1994).[16] Because these are often rated on Likert scales of

---

[16] For instance, McCarty and Shrum (1994) mixed in such terms when discussing the "value factor" of collectivism, which one might alternatively (we claim) think of as the value orientation "collectives over individuals" or the hypothesis "we are better off promoting the collective rather than individuals." They occasionally suggested that both values and value orientations entail, imply, involve, or perhaps may even *be* goals (see previous discussion) and/or beliefs. (E.g., "By definition, values are enduring beliefs and fairly resistant to change"; p. 61.) Scott and Willits

importance, agreement, or behavioral frequency, we believe that subjects might implicitly propositionalize them during these ratings. For instance, asking subjects to rate, from their perspectives, the importance of "warm relationships with others" is equivalent, in our view, to asking for a rating of the believability of the hypothesis, "Warm relationships with others are very important." Again, in our view, we would expect people to also implicitly consider the alternative, "Warm relationships with others are rather *un*important"—as well as a host of evidence (e.g., some "is" statements) and other hypotheses (e.g., some "ought" statements) that help to support or weaken these propositions.

Future Work and Future Worlds

The Biggest (Social) Problem: Another Instance of Satisfying Constraints

Releases of the Convince Me software are donated to a biology education package (e.g., Schank et al., 1995). But because the system is domain independent, it should be just as applicable to a host of other fields. In thinking of other potential applications, we have tried to focus on the societal good that this reasoner's workbench system might achieve, given our belief that the system is relevant to clarifying how people reason about social problems. More generally, if the system were truly useful, what would be the most important issue that one could address with it? We have actually put this type of question to a number of colleagues and students. The reader is invited to think of your own answer to this question, which we usually word as: What is the biggest problem/challenge facing humankind (or the Earth) during the next 50–100 years?

While you think about the question, you might introspect about your reasoning. Do you leap to a response, perhaps having had thoughts about this before? If so, perhaps you find yourself recalling the response's rationale, generating the argument's structure as you flesh out your memory. However, if you are like most people, you probably have not considered this specific question, and you generate a constraint-based argument from (a) various "big problems" that you

---

similarly added to this "word soup" as "ideas," "opinions," and "attitudes" are tossed in. (We thank Florian Kaiser for various discussions about these terms.)

have previously noted, (b) new alternative problems that you have just generated, (c) data and hypotheses that support these problems (which are themselves hypotheses) as candidates for "the biggest" problem, and (d) inhibitory relations among various reasons (and, of course, among the problems themselves—because not all can be "the" biggest). In essence, to answer this question, one carries out constraint-based reasoning (which probably involves some social components); this is precisely what Convince Me is designed to facilitate and explicate as it helps people epistemically categorize their propositions and clarify the relations among them. Thus, even the exercise of finding noteworthy problems, an unusually high-level activity, lends itself to analyses involving evidence, hypotheses, and (positive and negative) constraints. Indeed, a constraint-based system like *Convince Me* might not only help one uncover some of the relationships among various different kinds of problems (described later), it could also be used to explore what the possible trade-offs and synergies would be in addressing different problems.

To date, we have been intrigued by the variety of answers to the "biggest problem" question, although we note that the precise wording (and perhaps the timing and place) of the query may be critical. For instance, polls just prior to our questioning showed that the number one issue for Americans was crime (e.g., over the economy), yet almost none of our subjects answered our question with such responses.[17] Similarly, virtually none of our respondents listed "nuclear war," which would have been most topical less than 10 years ago (and may soon return to prominence). A sample of conference delegates at a recent international symposium in India (comprised of both Indian and other cognitive scientists) yielded proportions that were similar to those found among academic colleagues in the United States: A majority of the respondents chose a category of rather directly resource-related problems, which humankind might (at some point) be able to "fix."

---

[17] Indeed, a poll of 807 registered voters in our state, taken during late January of 1996, showed "immigration" as rivaling "economy/jobs" as the "single most important issue facing California" ("Examiner Poll," 1996). It is unlikely that many other (e.g., nonborder or recession-free) states would have yielded this pattern.

Overpopulation appeared to be the modal response (both in the United States and among the international delegates), which the authors would have predicted.[18] But others in this category saw overpopulation as merely a symptom of a more fundamental problem, namely that our resources are poorly distributed, or are either being depleted, have been depleted, or are being used to damage our environment during such depletion.[19] (Thus overpopulation is inhibited for these respondents by alternative, constraint-driven, complexes of data and theory.)

The second category of responses take a more mentality-related approach—in essence, inhibiting responses from the first category. For such respondents (although some subjects offered responses from both categories), they saw the tendency toward intolerance, greed, out-group exclusion, nationalism, not enough love and sharing, and so forth, as the major culprit or culprits. These problems seem less likely to have obvious solutions, although some might argue that following the Golden Rule is in some ways easier than getting people to reduce the number of babies they generate.

Depending on the particular demographics of the sample of people queried, the proportions of responses in the two main categories may differ widely.[20] An undergraduate cognitive science

---

[18] Even industry leaders like Steven Jobs see this as the biggest problem facing humanity (Wolf, 1996).

[19] In concert with such attitudes, Scott and Willits (1994) wrote about part of their 3,500-respondent 1990 Pennsylvania survey: "Only 45% felt that we are approaching the limit of the number of people the earth can support" (p. 245). (Note that various experts report that, although the planet's human population may never double beyond the current number, we will still have too many people for the resources available.)

[20] Beyond these two categories, responses are generally too ambiguous, vague, or all inclusive to fit well into either the resource- or mentality-related categories. These include uncertainty, lack of education (e.g., due to a lack of resources, or the undervaluing of education), the vague blaming

class yielded equivalent proportions of about .40 apiece, whereas an arguably more heterogeneous E-mail query (of primarily professionals, many with postgraduate degrees) yielded a majority of the mentality-related responses (e.g., highlighting greed).

One might see these categories of responses as rather mutually exclusive. After all, can there be more than one biggest problem? Still, if a decision maker (perhaps one in government) is having difficulty determining priority among such candidate problems, Convince Me might be able to help generate a kind of triage. For instance, one analyst might believe that overpopulation is a very strong candidate for biggest problem, say, in contrast to either resource management (a competitor from the first category) or "becoming better people" (from the second category). A rationale for this might include the reasoning that neither the best of intentions nor the best of technology and conservation will allow an unlimited number of humans to survive (e.g., given that they would obviously ultimately have only substarvational nutrition/energy resources). On the other hand, an analyst with a resource-management view might point out that, even if only two human beings populated the entire Earth, they still might destroy or eliminate life as we know it—for instance, if each had a million nuclear warheads (targeted around the globe) that were about to fire; in this case, overpopulation pales in comparison to resource management. This same scenario could also allow a third analyst to show that too much malice (e.g., expressed through warfare) can outweigh population and resource usage as problems, especially because one need not eliminate all resources with a multitude of nuclear weapons in order to cause human extinction. Of course, each of the arguments for these three competing "biggest problem" hypotheses could be increasingly more articulated in Convince Me. Determining the biggest problem is obviously a very complex endeavor, and involves difficult trade–offs, such as between the quality and quantity of life, between equity and sustainability in resource usage, and among various other constructs. Such

---

of groups ("governments," "people," etc.), and even that there is no consensus about what the biggest problem is, which leads to inaction.

trade–offs are also well suited to constraint-satisfaction analyses, for example, those embodied in Convince Me and its ECHO model.

Further Avenues of Practical Application

So, are the problems named as biggest by our respondents best considered as social or scientific ones? Again, the distinction seems somewhat irrelevant. Birth control is both a scientific and a social problem, as is resource management, pollution, global warming, and the destruction of rainforests. Similarly, from the mentality-related category, sharing what we have, understanding others, fostering equity—these are also areas in which both the scientific (or science-informed policy) and the social meet.

What, then, must we do? Specifically, what can cognitive scientists and social psychologists interested in constraint-based reasoning do to best increase the social good? Is the writing of archival articles enough? We think not, yet are conflicted as to how to proceed (cf. Kempton, Darley, & Stern, 1992, regarding environmental issues).

One highly attractive, socially responsible, approach would be to widely publicize what the biggest problems are and help address them. However, because few of us actually study these content domains, several drawbacks pop up. For instance, it might make our "day-job research" seem less significant, and it would take a long time to become an expert about these problems. Further, who would listen—and how might one best educate the populace and policymakers? Not only does it take considerable hubris to assume that one can "change things," it takes quite a bit of hubris to even assume that one's analyses and advocacies are correct on such grandiose issues!

For both analysis and policy, we are unfortunately largely reasoning under uncertainty (Tversky & Kahneman, 1974, etc.) in a world that does not provide its own natural control groups. Let us assume, for the sake of argument, that overpopulation is indeed the biggest problem. It may be that doing nothing would actually be best, ultimately, for the human race. Perhaps allowing a population-biologist's nightmare of overshoot-and-catastrophe to occur among humans will actually result in a much better (or even more numerous) subsequent society. On the other hand, slowing population growth now might mean we can successfully "build down" the population to

manageable levels before our collective health and welfare diminishes. Constraint-based systems can help with such decision making, although they certainly cannot fully eliminate the uncertainty involved.

As advocates for whichever of these two future-world models one chooses, one must still decide on the tactics to comprise a strategy for addressing the problem. One might try a direct approach, such as that of advertising (e.g., birth control, morality, recycling), or one might try a (seemingly) less direct approach, such as employing governmental rewards and punishments for desired and undesired behaviors. Still, with problems of such complexity and such temporal dynamics, assessments seem more like prognoses and solutions seem more like treatments. Constraint-based systems such as Convince Me can offer tactical help as well, but it remains to be seen whether they can adequately "scale up" to the complexity required by these conundrums (even at some reasonably intermediate level of abstraction).

Summary and Conclusions

The theory of explanatory coherence, first employed to model scientific reasoning, has been extended to the realm of social (and societal) reasoning. We suggest that no truly new features are required for this extension, indicating that scientific and social reasoning processes are significantly different primarily in domain, but not in kind. Hypotheses (or theories) and evidence (or facts) are employed in both domains, and even seemingly sociospecific notions such as values, goals, and candidate actions can be plausibly captured with evidence and hypotheses and some excitatory and inhibitory relations expressed among them. Further muddling the scientific/social distinction, scientific reasoning itself often involves social aspects.

We presented model-driven analyses and studies involving socially oriented reasoning (e.g., regarding the abortion debate, HIV quarantines, and marijuana legalization) to illustrate these points. In addition, a software environment, Convince Me, is offered as a reasoner's workbench that helps individuals articulate their thinking about such controversies. The system uses the connectionist model's output as feedback regarding the reasoner's articulation and coherence, suggesting ways in which constraint-satisfaction models may enhance the social good.

As reflected in our subjects' considerable concerns regarding social and scientific (or even technological) "biggest problems," systems that better describe, predict, and foster coherent reasoning will be in increasingly greater demand as the attendant complexities increase. Highlighting such concerns is a poem embedded in a "Calvin and Hobbes" cartoon (Watterson, 1992) in which an alien spacecraft that is plundering our planet's resources provides a metaphor for our considerations regarding the stewardship of the Earth. As the humans complain of the drastically depleted air and water, the alien ship broadcasts: "We're sorry to learn that you soon will be dead, But though you may find this slightly macabre, We prefer your extinction to the loss of our job."

It is likely that our future as a species will hold more (and presumably more important) socioscientific choices than the number of choices we have faced heretofore. Constraint-based models, especially embedded within knowledge-articulation systems such as Convince Me, offer much hope that we can make *better* choices. Therefore, continued research on this class of models seems vital—because the dilemmas represented by our many choices will certainly not be going away.

Acknowledgments

We thank Christine Diehl, Sergio Castro, Stephen Read, Lynn Miller, Christopher Ritter, Florian Kaiser, Elijah Millgram, Paul Thagard, Todd Shimoda, Steve Adams, Bob Branstrom, Larry Hamel, Michelle Million, George Lakoff, Phil Vahey, Rachel Ranney, the Reasoning Group, and others for their help and comments. Naturally, their assistance does not imply that they endorse each of the ideas expressed herein. We are also grateful to the Spencer Foundation and the Committee on Research of the University of California for funding some of these studies.

References

Bereiter, C. (1991). Implications of connectionism for thinking about rules. *Educational Researcher, 20*, 10-16.

Diehl, C. (1995, June). *Pragmatic and conceptual attributes of representational tools influence students' reasoning strategies*. Paper presented at the annual meeting of the American Psychological Society, New York.

Diehl, C., Castro, S., & Ranney, M. (1997, March). *Student models of hypothesis and evidence*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Diehl, C., Ranney, M., & Schank, P. (1995, April). *Multiple representations for improving scientific thinking* (Report No. TP-024-671). Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 392 842)

Einstein, A. (1950). Out of My Later Years. New York: Philosophical Library.

Examiner poll. (1996, February 4). *San Francisco Examiner*, p. A-12

Fiske, S., & Taylor, S. (1984). *Social cognition*. New York: Random House.

Fletcher, G. J. O. (1993). The scientific credibility of commonsense psychology. In K. H. Craik, R. Hogan, & R. N. Wolfe (Eds.), *Fifty years of personality psychology* (pp. 251-268). New York: Plenum.

Gabrys, G. (1989). HEIDER: A simulation of attitude consistency and attitude change. In S. Ohlsson (Ed.), *Aspects of cognitive conflict and cognitive change* (Tech. Rep. No. KUL-89-04). Pittsburgh: University of Pittsburgh, Learning Research and Development Center.

Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review, 98*, 254-267.

Gollwitzer, P. M., & Bargh, J. A. (Eds.). (1996).*The psychology of action: Linking cognition and motivation to behavior*. New York: Guilford.

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

Jensen, A. R. (1980). *Bias in mental testing*. New York: The Free Press.

Kahle, L. R. (Ed.) (1983). *Social values and social change: Adaptation to life in America*. New York: Praeger.

Kaiser, F. G., Ranney, M., Hartig, T., & Bowler, P. A. (1997). *Ecological behavior, environmental attitudes, and feelings of responsibility for the environment*. Manuscript in preparation, University of California, Berkeley.

Kempton, W., Darley, J. M., & Stern, P. C. (1992). Psychological research for the new energy problems: Strategies and opportunities. *American Psychologist, 47*, 1213-1223.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review, 103*, 284-308.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Lave, J. (1988). *Cognition in practice: Mind, mathematics, and culture in everyday life*. New York: Cambridge University Press.

McCarty, J. A., & Shrum, L. J. (1994). The recycling of solid wastes: Personal values, value orientations, and attitudes about recycling as antecedents of recycling behavior. *Journal of Business Research, 30*, 53-62.

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: Lawrence Erlbaum Associates.

Millgram, E. (1995). Inhaltsreiche ethische Begriffe und die Unterscheidung zwischen Tatsachen und Werten [Thick ethical concepts and the fact-value distinction]. In C. Fehige and G. Meggle (Eds.), *Zum moralischen Denken* (pp. 354-388). Frankfurt: Suhrkamp.

Millgram, E., & Thagard, P. (1996). Deliberative coherence. *Synthese, 108,* 63-88.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.

Norman, D. A. (1993). *Things that make us smart: Defending human attributes in the age of the machine*. Reading, MA: Addison-Wesley.

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.

Perkins, D. (1995). *Outsmarting IQ: The emerging science of learnable intelligence*. New York: Free Press.

Pervin, L. A. (Ed.) (1989). *Goal concepts in personality and social psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Putnam, H. (1981). *Reason, truth, and history*. Cambridge, England: Cambridge University Press.

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. (1996). Individual-centered vs. model-centered approaches to consistency: A dimension for considering human rationality. *VIVEK, A Quarterly in Artificial Intelligence*, *9* (2), 35-43.

Ranney, M., & Schank, P. (1995). Protocol modeling, bifurcation/bootstrapping, and *Convince Me*: Computer-based methods for studying beliefs and their revision. *Behavior Research Methods, Instruments and Computers, 27*, 239-243.

Ranney, M., Schank, P., Hoadley, C., & Neff, J. (1996). "I know one when I see one": How (much) do hypotheses differ from evidence? In R. Fidel, B. H. Kwasnik, C. Beghtol, & P. Smith (Eds.), *Advances in classification research:* Vol 5. (ASIS Monograph Series; pp. 141-158), 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). Artificial Intelligence in Education Society.

Ranney, M., Schank, P., & Ritter, C. (1992, January). *Studies of explanatory coherence using text, discourse, and verbal protocols*. Paper presented at the Third Annual Winter Text Conference, Jackson, WY.

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: Lawrence Erlbaum Associates.

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

Read, S. J., & Miller, L. C. (1993). Rapist or "regular guy"?: Explanatory coherence in the construction of mental models of others. *Personality and Social Psychology Bulletin, 19*, 526-541.

Read, S. J., Vanman, E. J., & Miller, L. C. (in press). Connectionism, parallel constraint satisfaction processes, and Gestalt principles: (Re)Introducing cognitive dynamics to social psychology. *Review of Personality and Social Psychology*.

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

Rumelhart, D., McClelland, J., & the PDP Research Group. (1986). Parallel distributed processing: Explorations in the microstructure of cognition (Vols. 1 & 2). Cambridge, MA: MIT Press.

Schank, P. (1995). *Computational tools for modeling and aiding reasoning: Assessing and applying the Theory of Explanatory Coherence*. Doctoral dissertation, University of California, Berkeley. (University Microfilms No. 9621352)

Schank, P., & Ranney, M. (1991). The psychological fidelity of ECHO: Modeling an experimental study of explanatory coherence. In J. K. Kruschke (Ed.), *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* (pp. 892-897). Hillsdale, NJ: Lawrence Erlbaum Associates.

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* (pp. 599-604). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Schank, P., Ranney, M., & Hoadley, C. (1995). *Convince Me* [Computer program and manual]. In J. R. Jungck, N. Peterson, & J. N. Calley (Eds.), *The BioQUEST library*. College Park, MD: Academic Software Development Group, University of Maryland.

Shultz, T. R., & Lepper, M. R. (1992). A constraint satisfaction model of cognitive dissonance phenomena. In K. J. Hammond & D. Gentner (Eds.), *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* (pp. 462-467). Hillsdale, NJ: Lawrence Erlbaum Associates.

Shultz, T. R., & Lepper, M. R. (1996). Cognitive dissonance reduction as constraint satisfaction. *Psychological Review, 103,* 219-240.

Scott, D., & Willits, F. K. (1994). Environmental attitudes and behavior: A Pennsylvania survey. *Environment and Behavior, 26*, 239-260.

Simon, H. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*, 99-118.

Spellman, B. A., Ullman, J. B., & Holyoak, K. J. (1993). A coherence model of cognitive consistency: Dynamics of attitude change during the Persian Gulf War. *Journal of Social Issues, 49* (4), 147-165.

Taylor, J. (1985). Mona. On *That's why I'm here*. [LP] New York: Country Road Music (BMI).

Tetlock, P. E. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly, 48*, 227-236.

Tetlock, P. E. (1992). The impact of accountability on judgment and choice: Toward a social contingency model. *Advances in Experimental Social Psychology, 25*, 331-376.

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

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

Thagard, P. R. (1996). *Mind: Introduction to cognitive science*. Cambridge, MA: MIT 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: AAAI Press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124-1131.

Watterson, B. (1992). *Calvin and Hobbes* [cartoon]. Kansas City, MO: Andrews and McMeel, Universal Press Syndicate.

Wertheimer, M. (1982). *Productive thinking* (Pheonix ed.). Chicago: University of Chicago Press.

Wolf, G. (1996, February). [Interview with Steven Jobs]. Steve Jobs: The next insanely great thing. *Wired, 4* (2), pp. 102-107, 158, 160, 162-163.