Explanatory coherence in the construction of mental models of others

Stephen J. Read
Psychology Department
University of Southern California
Los Angeles, CA 90089-1061
BITNET: READ@USCVM

Lynn Carol Miller
Communication Arts and Sciences
University of Southern California
Los Angeles, CA 90089-1694
BITNET: LMiller@USCVM

Abstract

A unified model of social perception, integrating causal reasoning and impression formation (Miller & Read, 1991), provides an account of how people arrive at coherent representations of others and explain their behavior. The model integrates work on a knowledge structure approach (Schank & Abelson, 1977) with Kintsch's (1988) construction-integration model and Thagard's (1989) model of explanatory coherence. We explore two issues in social perception. First, we show how the model can be used to explain trait inferences, where traits are treated as frames, composed of goals, plans, resources and beliefs. Second, we examine how people might combine inconsistent traits to arrive at a coherent model of another, an example of conceptual combination.

Introduction

In everyday social interaction we must process a constant stream of information about others. How do we transform this complex array of behavior and inference into a coherent model of others? To address this question we propose the following model, based on four fundamental assumptions. First, individuals make sense of everyday social interactions by creating a coherent scenario or story from the sequences of actions they observe (e.g., Abelson & Lalljee, 1988; Pennington & Hastie, 1986; Read, 1987). They do so by making inferences about the cause-effect relations among the behaviors, the actors' goals, and various higher order structures, such as themes that characterize it. Second, the development of these scenarios is guided by principles of explanatory coherence, such as simplicity and breadth of explanation (Thagard, 1989). The greater the perceived coherence of the scenario, the more apt people are to feel that they understand events. Although little social perception theory or research addresses how people construct representations of causally connected social behaviors (for an exception see Pennington & Hastie, 1986), this problem is central to the current approach. Third, the meanings of social behaviors is dynamic; the result of mutual influence among cognitive elements currently active in the system (Miller, Bettencourt, DeBro, & Hoffman, 1993). Historically, Asch (1946) and Heider (1958) argued that such gestalt processes played a central role in social perception. Happily, recent work on parallel constraint satisfaction processes in connectionist modeling now provide a concrete computational implementation of Gestalt-like processes in the interpretation of simultaneously constrained cognitive elements. Fourth, making social inferences and understanding social interaction depends on extensive physical and social knowledge (Abelson & Lalljee, 1988; Heider, 1958; Miller & Read, 1987, 1991; Read, 1987; Schank & Abelson, 1977). Thus, we emphasize the role of concrete knowledge in explanation, in contrast to the emphasis on abstract, logical analysis in many other models of attribution (Cheng & Novick, 1992; Jones & Davis, 1965).

Model

Developing social explanations, we argue, involves two steps. First, input activates concepts somewhat "promiscuously" (Kintsch, 1988) through a spreading activation process (e.g., Collins & Loftus, 1975). Initially, there is little check on the consistency of concepts with each other. Thus, several alternative explanations of the same event may be concurrently activated. Concepts are linked in a heterogeneous network, with a preference for connecting concepts that have causal and intentional relations (Galambos, Abelson, & Black, 1986). At this stage the concepts may be relevant, irrelevant, or even inconsistent with the eventual explanation of the event. There are three ways in which concepts can be linked. First, they may be positively linked so that the activation of one concept increases the activation of another (e.g., goal or causally related concepts). Second, they may be negatively linked, as when two concepts are inconsistent or contradict each other. Here the activation of one concept decreases the activation of another.

Third, concepts may be unlinked.

Once the initial network is built, a coherent representation is constructed by applying a parallel constraint satisfaction process to the network (e.g., Kintsch, 1988; Rumelhart, McClelland, and the PDP research group, 1986; Thagard, 1989) that implements Thagard's (1989) model of Explanatory Coherence. This iteratively converges on a pattern of activation that represents the best "compromise" among the constraints imposed by the links among the nodes. Because the activation of a proposition indicates its degree of acceptability, the degree to which the individual believes that the proposition describes the world, highly activated concepts are treated as the representation of the interaction up to that point. A new processing cycle is initiated for each action in a social interaction: a new network is built in working memory -- consisting of new input, newly activated associated concepts, and information passed along from the previous cycle. Inferences, such as higher order structures, that received high levels of activation will be among those concepts that are passed on from the previous cycle. The explanatory coherence of this network is then evaluated. Nodes with high levels of activation are added to the representation in long term memory. As new input is received, more inferences are made, and broader structures are built.

Although Thagard (1989) has implemented his theory in a connectionist computer program (ECHO), its principles can be considered separately from its computational implementation. However, as we describe the principles we will also explain how they are implemented in ECHO. The principles include (a) *breadth* – an explanation that explains more facts is more coherent. In Figure 1, Trait 3 (T3) is a better explanation than either Trait 1 (T1) or Trait 2 (T2) because it explains more and thus receives more activation. (b) *Parsimony or simplicity* – the explanation requiring fewer hypotheses will be more coherent, because the activation from each fact is divided among its explanations. In Figure 1, we have multiple facts to be explained and either one or several explanatory hypotheses that explain them. Trait 3 is a better explanation, because the other explanation requires both traits together to explain both behaviors. (c) *"Being explained"* --explanations are better if they can, in turn, be explained, because explanations send activation to what they explain. A trait can be explained by such factors as the previous history of the individual, such as socialization, major life events, or genetics. (d) *"Unexplained data"* -- the goodness of a trait as an explanation should be reduced if some behavior is unexplained. This is implemented by increasing the decay rate for activation in proportion to the amount of unexplained evidence. (e)*"Unification,"* -- a set of explanatory hypotheses is more coherent when they *jointly* explain *all* the evidence, without requiring unique explanatory hypotheses to explain some pieces of evidence. (f) Explanations supported by an *analogy* to another system with the same causal structure should be more coherent because the analogous explanation provides activation to the explanatory hypotheses.

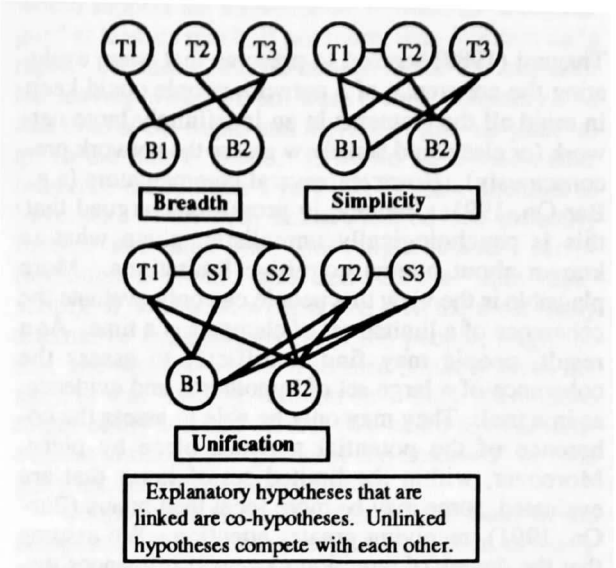

Figure 1. Graphic examples of some of the principles of explanatory coherence

Finally, the coherence of explanations is comparative, being a function of the coherence of alternatives. Thus, the acceptability of a mediocre explanation will decrease when a better alternative is available. This follows because of the inhibitory links among competing explanations.

Read and his colleagues (Read & Cesa, 1991; Read & Marcus-Newhall, in press) have provided experimental evidence for most of these principles in the construction of social explanations. Schank and Ranney (1991) have provided evidence for some of these principles in thinking about the behavior of physical systems.

In this model, trade-offs among principles are done implicitly, through the summation of activation. For example, suppose we have two alternative explanations for a set of data, one consisting of two explanatory hypotheses and the other of a single hypothesis. Because the explanations compete, they will have an inhibitory relation. If both explanations explain the same two pieces of evidence, then the simpler explanation should win, because it receives more activation, not having to divide the activation from the facts. But if we gather new evidence that can only be explained by the broader explanation, at some point, as the amount explained by the broader explanation increases, the broader explanation will receive higher activation then the simpler explanation and suppress it. Other principles similarly trade off.

Attentional focus

Thagard (1989) seemed to presume that when evaluating the coherence of a network people could keep in mind all the elements in an indefinitely large network (or else could somehow assess the network preconsciously). However, several commentators (e.g., Bar-On, 1991; Ranney, in press) have argued that this is psychologically unrealistic, given what is known about human cognitive limitations. More plausible is the view that people can only evaluate the coherence of a limited set of elements at a time. As a result, people may find it difficult to assess the coherence of a large set of hypotheses and evidence, as in a trial: They may only be able to assess the coherence of the potential network piece by piece. Moreover, within the limited set of items that are evaluated, some may be more focal than others (Bar-On, 1991), receiving greater attention. We assume that the degree of attention to a node influences the amount of activation it can send and receive. Nodes receiving higher levels of attention can both give and receive more activation.

Applications of the Model

To provide a concrete basis for our applications of the model consider the example of the William Kennedy Smith rape trial.

> William Kennedy Smith met Patricia Bowman at a West Palm Beach night spot. They danced, talked, and then left for the Kennedy family's compound. What followed, Patricia claimed, was a vicious rape; But William claimed it was consensual sex. Was William -- as his defense attorney argued -- simply a gregarious, "regular guy" attacked by a "woman scorned"? Or, was he a master manipulator attacking an unsuspecting, caring mother?

Over time, as new testimony was introduced, people might have entertained quite different "models" of William. At one time, he may have been viewed as "manipulative;" at another time, he may have been seen as "a regular guy." Some may have perceived Patricia as "revengeful." In contrast, the prosecution presented Patricia as a "caring mother."

But how do individuals "get up to" and choose among such trait and person inferences given the array of information provided? Various principles of coherence, together with the "evidence" and inferences available at different times might have supported each of these "models" of William and Patricia. But in addition, to fully address this question, we ultimately need to better understand the "deep structure" of person assignations such as traits. What, for example, are the components of traits, the "slots" that must be filled in specific ways to support

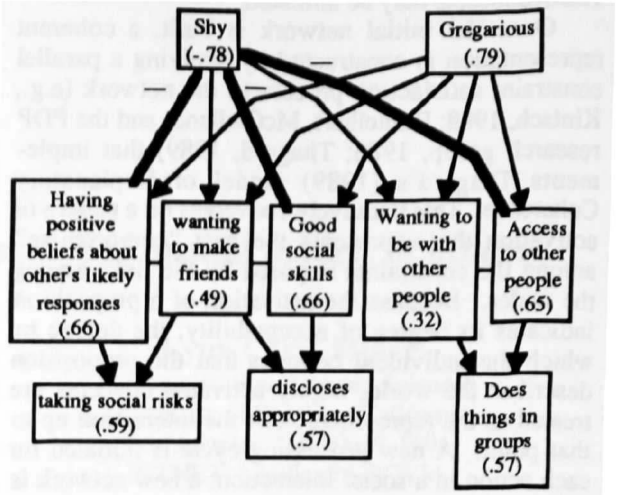

Figure 2. Explanatory coherence of two competing traits with multiple goals in common, but different beliefs and resources.

one trait attribution over an alternative? Furthermore, when information is already organized around traits, what role might the deep structure of each trait play in the way multiple traits are combined? Below we consider the possible deep structure of traits and the making of trait inferences

Trait inferences

We have proposed (Miller & Read, 1987; Read & Miller, 1989) that traits are chronic configurations of an individual's goals, plans, resources, and beliefs. For instance, an individual with the following chronic set of goals (e.g., wanting to make friends and wanting to be with people), plans (e.g., doing things with groups, taking social risks), resources (e.g., social skills, access to people, good encoder of nonverbal behavior), and beliefs (e.g., people are fun and rewarding), is more apt to be described as "gregarious" than someone with the same chronic goals but with different plans, beliefs, and resources. Thus, Figure 2, which presents the activations resulting from an ECHO run on this configuration of information, suggests that although individuals who are shy may share the same goals as those who are gregarious, "gregarious" is more likely to be activated when other components of the trait, such as beliefs and resources that are consistent with gregarious, but contradict shy, are activated.

Inferences about *goals* should be particularly important to making trait inferences (Miller & Read (1987; 1991; also see Jones & Davis, 1965). Recently, Read, Jones and Miller (1990) demonstrated that this is the case for a set of interpersonal traits. Behaviors that were most prototypical of a

trait, such as gregarious, were also viewed as most likely to achieve the goals ("wanting to make friends" and "wanting to be with people") that were strongly associated with that trait. Furthermore, behaviors that were more highly goal related led to greater confidence about the trait. Of what use might the centrality of goals in trait inferences be in understanding inferences about William? Interestingly, many of William's behaviors are potentially associated with "gregarious": going to the bar with his uncle, Senator Ted Kennedy, and his cousin, talking to people at the bar, dancing, listening to what others say, etc. As Read et al.'s (1990) findings suggest, we are likely to infer that William enacted these behaviors because he wanted to make friends, or wanted to be with others. With no other information, this might have been a reasonable "default" inference, not only for the jurors, but also for Patricia. But these same behaviors could have been in the service of an alternative goal, such as wanting to manipulate Patricia, and there was testimony that William's behavior changed, becoming much more aggressive after he arrived at the family estate. This behavior, in combination with the earlier bar behavior, could more coherently be described as "manipulative." This follows from the principle of breadth, because "manipulative" could explain the set of behaviors at the bar, as well as William's behavior after they left the bar, as employed to take advantage of Patricia. And complaints from other women about similar interactions with William (analogy), might provide further activation to manipulative (at least for those exposed to this information). Thus, manipulative would provide a more coherent explanation of William's various behaviors than gregarious.

Miller and Read (1987; 1991; Read & Miller, 1989) have further argued that oftentimes a configuration of goals underlies a given trait. For instance, two goals for gregarious are "wanting to make friends" and "wanting to be with people". When an actor performs behaviors strongly related to both goals, rather than to just one goal, individuals should make more confident trait inferences. The results of Read, Jones, and Miller (1990) support this expectation. This is consistent with the principles of breadth and unexplained data, in that as a trait can explain more goals and leaves less evidence unexplained, this trait should be more acceptable. In addition to goals, other structures, such as plans, resources, and beliefs are important for making inferences about behavior. For instance, consider the behaviors in the bar such as "dancing with William", and "talking to William" in the early morning hours, and then going with William to a private, isolated patch of beach near the Kennedy compound. These behaviors and resources (e.g., the car) are likely to activate many of the components (e.g., goals, plans, expectations) of a typical "one night stand script" (Miller, et al., 1993). This script would support the inference of William as "a regular guy" at least as much, if not more, than William as "a rapist" because Patricia's actions at the bar and later her leaving with William were clearly voluntary. At that time of the night, and leaving a "pick up joint" to go to the man's "place," it would be a reasonable inference on William's part that Patricia intended to have sex. That both acknowledged sex probably further activates the "one night stand" script (although it should also activate the "date rape" script). If she hadn't wanted sex, then the most likely alternative explanations for some people, might be that she was naive (which seemed unlikely given her age), or was "leading William on". Such attributions might easily activate others. We heard several individuals remark, "she got what she deserved" or "she got what she should have expected".

Scripts, such as the one above, have many of the components that we have argued are typical of many traits. In fact, many traits seem to have a "story structure" involving goals, plans, resources and beliefs (Miller & Read, 1991). For example, the trait revengeful, that was associated with Patricia by the defense, has embedded in it the following story: An individual feels that another did her a wrong and she had the resources with which to hurt that individual and wishes to do so or did so. Such event scripts can be thought of as a general "frame" possessing a number of slots which can be filled by the appropriate concept. For revengeful the slots are (1) a *behavior*, consisting of bringing charges against William, (2) *consequences of the behavior* such as hurting William's reputation, sending him to jail or hurting him professionally, (3) *the roles* (Patricia was a "woman scorned"; William was a "regular guy" who acted out a typical sexual script for a "one night stand") and the characteristics of the participants (Patricia was emotionally unstable; because of that it only took being angry with William for using the wrong name after sex to provoke Patricia to try to hurt William; she was manipulative; William was gregarious, although sleazy), (4) *resources* involved (she gathered evidence such as goods from the compound to make a case that she was there that night; she provided enough evidence to the prosecution that they pursued the case in court; access to the media could also serve to make a case against William), and (5) the *goals* and intentions of the participants (Patricia's goal was to hurt William to make him pay for having hurt her). If these components receive greater activation than alternatives, "revengeful" should be most activated: Thus, it was critical for the defense to activate these "slots" in building their case.

Trait combinations

Clearly, people are characterized by a host of charac-

teristics. How do people combine these different characteristics to arrive at coherent images of others? This is an example of the problem of conceptual combinations. Several researchers (e.g., Asch & Zukier, 1984; Hastie, Schroeder, & Weber, 1990; Kunda, Miller, & Claire, 1990) have suggested that by examining how people combine discordant information (e.g., about roles or traits), processes, such as causal reasoning, that are implicit in such resolutions are exposed. Asch and Zukier (1984) examined the resolutions when targets were characterized by discordant trait pairs such as cheerful-gloomy or generous-vindictive. Although subjects found it relatively easy to arrive at coherent interpretations of these targets, the resolutions were not averages of trait evaluations. Consider generous-vindictive. Many people suggested that the individual was only apparently generous, but was actually using his/her apparent generosity in gaining revenge. Here vindictive strongly modifies our interpretation of the individual's generosity.

This resolution process can be analyzed as follows (for a partially related model of conceptual combinations of social concepts see Hastie et al. (1990)). Traits are frames consisting of slots, including goals, plans, resources, and beliefs. Information is stored about the default for these slots and the range of concepts that can fill them. Thus, generous and vindictive have, as part of their representation, slots for the associated goals and slots for the behaviors that can achieve those goals. Further, there are constraints on the values that can fill those slots. For instance, the range of goals for vindictive seems fairly narrow. It is hard to think of vindictive individuals as having any goal other than to hurt someone. In contrast, there seems a greater range for the goals of generous. Although the most likely goal is probably something like giving to someone, there are other reasons why someone is generous, including self-presentational and strategic reasons. The tasks in such resolutions suggest that a model that assumes a parallel constraint satisfaction process may be particularly useful. Each trait has multiple slots, providing numerous ways in which two traits can be related. Further, there is a range of potential values for each slot, with constraints on these values. In addition, there are constraints on which slots can be plausibly related. Therefore, any successful resolution of two discordant traits requires that one solve multiple sets of constraints and choose the resolution that does the "best" job of satisfying those constraints.

Asch and Zukier (1984) have identified a number of strategies; each seeming to depend upon a different combination of the components that we have argued are key components of traits. For instance, many subjects suggested that being intelligent enabled an individual to be witty. Here one trait, intelligence, provides a **resource** that enables the achievement of the **goal** associated with the other trait. One trait may also provide the means, plan or behavior to achieve the other. For example, many subjects said that an individual could be both strict and kind, because an individual, such as a parent or teacher might be strict so that a child performed well or behaved well and so developed better. Here strict can be seen as providing a **plan** for achieving the **goal** of the trait kind. Or, one trait can cause or lead to another. Many subjects said that an individual could be both hostile and dependent, because being dependent on others might lead to resentment and thus hostility. Here one trait instigates the **goal** associated with the other. Another strategy involves interpolation in which a characteristic that mediates a relationship between two traits is inferred. For example, if an individual is both intelligent and unambitious then subjects may infer that this individual's attempts have been unsuccessful, leading to discouragement and a lack of ambition. Subjects are inferring that the achievement **goals** typically associated with intelligence have been blocked so that the individual has developed the lack of a **goal** typical of being unambitious.

As we and others have noted (e.g., Read & Miller, in press; Schultz & Lepper, 1992; Spellman, Ullman, & Holyoak, in press), parallel constraint satisfaction processes capture the Gestalt-like processes that are central to many key theories in social psychology such as various consistency theories (e.g., Balance theory (Heider, 1958) and Cognitive Dissonance (Festinger, 1957)). Most work on social reasoning assumes a somewhat serial process, or else, vaguely refers to an unspecified parallel process. Models based on parallel constraint satisfaction processes could greatly expand how we think about social reasoning. Such processes allow for the explicit, dynamic, integration of multiple sources of constraint and allow for simultaneous tradeoffs among them.

References

Abelson, R. P., & Lalljee, M. (1988). Knowledge structures and causal explanation. In D. Hilton (Ed.), Contemporary science and natural explanation....(p. 175-203). London: Harvester Press

Asch, S.E.,& Zukier, H. (1984).Thinking about persons. Journal of Personality and Social Psychology, 46,1230-1240.

Bar-On, E.(1991). Locally-Coherent views. Draft October 1991, Department of Science Education, Technion, Israel.

Cheng, P.W., & Novick, L. R. (1992) Covariation in natural causal induction. Psychological Review, 99, 365-382.

Collins, A.,& Loftus, E. (1975).A spreading activation theory of semantic processing. Psycho-

logical Review, 82,407-428.

Festinger, L. (1957). A theory of cognitive dissonance. Evanston, IL: Row, Peterson.

Hastie, R., Schroeder, C., & Weber, R. (1990). Creating complex social conjunction categories from simple categories. Bulletin of the Psychonomic Society, 28, 242-247.

Heider, F. (1958). The psychology of interpersonal relations. New York: Wiley.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 2, pp. 219-267). New York, Academic Press.

Kintsch, W. (1988). The role of knowledge in discourse comprehension. Psychological Review, 95, 163-182.

Kunda, Z., Miller,D.T., & Claire, T. (1990). Combining social concepts: The role of causal reasoning. Cognitive Science, 14, 551-577.

Miller, L.C., Bettencourt, B. A., Debro, S., & Hoffman, V. (1993). Negotiating Safer Sex: A Dynamic Interpersonal Process. In J. Pryor and G. Reeder (Eds.). The Social Psychology of HIV Infection. Hillsdale, N.J.: Erlbaum.

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

Murphy, G. L., & Medin, D. L. (1985).The role of theories in conceptual coherence. Psychological Review, 92, 289-316.

Pennington, N. & Hastie, R. (1986). Evidence evaluation in complex decision making. Journal of Personality and Social Psychology, 51, 242-258.

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

Read, S. J. (1987). Constructing causal scenarios: A knowledge structure approach to causal reasoning. Journal of Personality and Social Psychology, 52, 288-302.

Read, S.J., & Cesa, I. L. (1991).This reminds me of the time when... Journal of Experimental Social Psychology, 27,1-25.

Read, S. J., Jones, D. K., & Miller, L. C. (1990). Traits as goal-based categories:.. Journal of Personality and Social Psychology, 58,1048-1061.

Read, S. J., & Marcus-Newhall, A. (in press). The role of explanatory coherence in the construction of social explanation. Journal of Personality and Social Psychology

Read, S. J., & Miller, L. C. (1989). Inter-personalism: Toward a goal-based theory of persons in relationships. In L. Pervin (Ed.), Goal concepts in personality and social psychology (pp. 413-472). Hillsdale, NJ: Erlbaum.

Read, S. J., & Miller, L. C. (in press). Dissonance and balance in belief systems: The promise of parallel constraint satisfaction processes. In R. C. Schank & E. J. Langer (Eds.). Beliefs, reasoning, and decision making: Psycho-logic in honor of Bob Abelson. Hillsdale, NJ: Erlbaum.

Rumelhart, D. E., & McClelland, J. L. (1986). Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1. Foundations. Cambridge, MA: MIT Press/Bradford Books.

Schank, R. C., & Abelson, R. P. (1977). Scripts, plans, goals and understanding. Hillsdale, NJ: Erlbaum

Schank, P.K., & Ranney, M. (1991).An empirical investigation of the psychological fidelity of ECHO. Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society, 892-897. Hillsdale, NJ: Erlbaum.

Spellman, B.A., Holyoak, K. J., & Ullman, J. (in press). Shifting views of the Persian Gulf War:..Journal of Social Issues.

Thagard, P. (1989). Explanatory Coherence. Behavioral and Brain Sciences, 12, 435-467.