A Feedback Neural Network Model of Causal Learning and Causal Reasoning

Stephen J. Read (read@rcf.usc.edu)
Department of Psychology, University of Southern California
Los Angeles, CA 90089-1061

Jorge A. Montoya (gmontoya@rcf.usc.edu)
Department of Psychology, University of Southern California
Los Angeles, CA 90089-1061

ABSTRACT

We present a feedback or recurrent, auto-associative model that captures several important aspects of causal learning and causal reasoning that cannot be handled by feedforward models. First, our model learns asymmetric relations between cause and effect, and can reason in both directions between cause and effect. As a result it can represent an important distinction in causal reasoning, that between necessary and sufficient causes. Second, it predicts cue competition among effects and provides a mechanism for them, something which can only be done with feedforward models by assuming that two separate networks are learned, a highly non parsimonious assumption. Finally, we show that contrary to previous claims, a feedforward model cannot handle Discounting and Augmenting in causal reasoning, although a feedback model can. The success of our feedback model argues for a greater focus on such models of causal learning and reasoning.

Introduction

Connectionist models of causal learning and reasoning have relied on feedforward networks (e.g., Gluck & Bower, 1988; Shanks, 1991; Van Overwalle, 1998). However, as we have recently shown, feedforward networks have serious limitations as models of causal learning and reasoning (Read & Montoya, in press). In that paper, we outlined an alternative, a feedback or recurrent model, that can handle phenomena that a feedforward model cannot. In the current paper we examine further implications of this kind of model for phenomena that feedforward models cannot handle, such as asymmetries in causal learning and reasoning, and cue competition for consequences or effects.

In previous work we have examined how this kind of model can handle a number of phenomena in causal learning and causal reasoning. Read and Montoya (in press) have demonstrated that it can successfully simulate many of the classic phenomena from the animal and human causal learning literature, such as blocking and conditioned inhibition, to which the Rescorla-Wagner model (Rescorla & Wagner, 1972) and feedforward models with delta-rule learning (e.g., Gluck & Bower, 1988; Shanks, 1991), have been applied. Read and Montoya also demonstrated that this auto-associative model, which is a parallel constraint satisfaction model, deals with the principles of explanatory coherence discussed by Thagard (1989, 1992) and experimentally demonstrated by Read and Marcus-Newhall (1993) and Read and Lincer-Hill (1998) (see also Ranney, in press; Schank & Ranney, 1991, 1992). Finally, several papers (Montoya & Read, 1998; Read & Miller, 1993) have shown that this kind of model can simulate the Discounting and Augmenting principles in causal reasoning (Kelley, 1971), as well as the role of factors, such as construct accessibility and causal strength, that may underlie the closely related Correspondence Bias or Fundamental Attribution Error (Jones, 1990; Ross, 1977).

In the current paper, we focus on the implications for causal learning and reasoning of a central aspect of this model: all nodes are completely interconnected, with an independent link going in each direction between each pair of nodes. This has three implications which we will examine. First, because each pair of nodes is joined by two links, one in each direction, it is possible to reason both from cause to effect and from effect to cause. In contrast, with the feedforward models previously investigated in causal learning and reasoning, it is only possible to learn and reason in one direction, typically from cause to effect. Second, because each member of the pair of links can have different strengths, the link from cause to effect can have a different strength than the link from effect to cause. As a result, with this model one can learn asymmetric relations between cause and effect, and use these asymmetric relations in causal reasoning. Third, because the network is totally interconnected, it can learn relations among possible causes of an event. In contrast, in the feedforward networks used in this domain the only links are forward, from cause to effect. It is not possible to learn links among causes. One implication of this, we will argue, is that the standard feedforward model is incapable of handling either discounting or augmenting in causal reasoning, whereas our model can handle both phenomena.

An Auto associative Model.

Our model is based on McClelland and Rumelhart's (1988) auto-associator, which is a single layer auto-associative network with all units completely interconnected. Each unit receives input from other nodes and simultaneously sends activation to other nodes. Because of the feedback relations, this network functions as a parallel constraint satisfaction system, acting to satisfy multiple simultaneous constraints among elements in the network. Links are modified by delta-

rule learning and each link in a pair can end up with a different weight. All of the nodes can receive input from both the environment and other nodes. Thus, both cause and effect nodes can be activated by environmental cues. (Although Thagard's ECHO model is also a feedback model, it assumes that both links between pairs of nodes have identical weights. Thus, there is no way to represent asymmetric causal relations in ECHO and no way to examine the role of differences in links from cause to effect and effect to cause. Further, because ECHO has no learning mechanism, it cannot learn causal links (however, see Wang, Johnson, and Zhang (1997) who have recently added delta rule learning to ECHO).)

This network can learn associations among all the elements that co-occur. That is, not only can it learn the relation between the effect X and potential causes A and B, it can also learn the association between the two potential causes. In contrast, in feedforward networks, there are links in only one direction, from input nodes to output nodes. Output nodes only receive activation from the input nodes, and cannot be directly activated by the environment. Also, there are no links among the nodes in a layer; the only links are between layers. Thus, it cannot learn associations between causes.

Processing in the auto associative network proceeds as follows. After input is received, all the units in the network are synchronously updated at each cycle by an activation function that is essentially the same as that employed in ECHO (Thagard, 1989; 1992) and in Rumelhart and McClelland's (1986) interactive-activation and competition model, as well as in a handful of other models they have explored. This activation function is:

$$a_j(t+1) = a_j(t)(1-d) + \begin{cases} net_j(max - a_j(t)) & \text{if } net_j > 0 \\ net_j(a_j(t) - min) & \text{if } net_j \le 0 \end{cases}$$

$$\text{where } net_j = (istr)[\Sigma w_{ji} a_i] + (estr)ext$$

The only minor difference in this activation function for the auto-associative architecture, compared to other models in which it has been used, is that the total input net_j is now determined by external input from the pattern vector *ext*, as well as the sum of weighted inputs from other units within the network with activations from the previous cycle, $\Sigma w_{ji} a_i$. Note that the internal input and the external input are scaled, by *istr* and *estr*, respectively.

After the system completes a number of processing cycles (defined by the user), the delta rule (or Widrow-Hoff rule) (Widrow & Hoff, 1960) is applied to the network to compare the external input pattern to the internal inputs to units. This learning regime reduces the difference between internal and external inputs to units, by modifying the weights among the nodes, so that the internal input comes to reproduce or match the external input to the units. Hence, the *desired* activation of a unit is determined by the set of external inputs to that unit. The discrepancy between the *desired* and *actual* activation of a unit is the measure of error used in delta rule learning. Weight change is given by:

$$\Delta weight_{ji} = lrate\ (t - a_j)\ a_i,$$

where lrate is the learning rate, t is the target or external activation, a_j is the internal or actual activation, and a_i is the activation of the node sending activation to a_j.

Learns and Uses Asymmetries in Causal Relations.

One advantage of this model is that separate links exist from cause to effect and from effect to cause. As a result, this model is able to learn any asymmetries that might exist in these relationships. Further, having learned these asymmetries, they can be used in causal reasoning.

In contrast, neither current associative models (e.g., Gluck & Bower, 1988; Shanks, 1991; Van Overwalle, 1998) nor Cheng's (Cheng & Novick, 1990, 1992) probabilistic contrast model can learn separate relations for cause to effect and effect to cause. In fact, both capture the relationship from cause to effect, but not the reverse relationship. Thus, these models cannot learn asymmetries in cause-effect and effect-cause relations. Further, these models do not allow for reasoning in both directions.

Several authors (e.g., Shanks, Lopez, Darby, & Dickinson, 1996) suggest that one could capture the two different directions of causal learning by using two feedforward networks, one with causes as inputs and the other with effects as inputs. However, with recurrent networks, such as the present model, only one network is required. This is much more parsimonious than assuming that an individual would require two separate networks to capture bi-directionality in causal learning and reasoning.

Table 1 gives a set of learning trials that result in asymmetric learning of links, such that cause A has a stronger forward link to X than does cause B, whereas effect X has a stronger backward link to cause B than to cause A. In this example, assume that we are learning and reasoning about possible causes of a forest fire (X). One possibility is lightning (A) while another is a campfire (B).

Table 1: Learning History for Asymmetry in Causes

Simulation	Unit	Learning history	Epochs
Asymmetry	A	+ + ...	20
	B	. . + + + + + + + +	
	X	+ + + + + + ...	

Because of the pattern of covariation, asymmetric causal relations are learned. The model learns that if it occurs, lightning is more likely to cause a forest fire, than is a campfire. However, it also learns that if there is a forest fire it was more likely preceded by a campfire then by lightning. This asymmetry is apparent in both the activations when causes and effects are separately tested and in the patterns of weights that are learned.

When we separately activate the two causes, A (lightning) alone leads to a higher activation for X (forest fire) than does B (campfire), .35 versus .16. However, if effect X alone (forest fire) is activated then cause B (campfire) is more highly activated, .37, than is cause A (lightning), .27.

The connection strengths leads to the same conclusion. The connection from A (lightning) to X (forest fire) is stronger than the connection from B (campfire) to X (forest fire), 1.58 versus .74. However, the connection from

X(forest fire) to B(campfire) is stronger than the connection from X(forest fire) to A (lightning), 1.88 versus .94. The model has learned that the occurrence of lightning is more likely to cause a forest fire than is the occurrence of a campfire. However, it has also learned that if a forest fire occurs that it is more likely to be caused by a campfire.

Such asymmetries seem to be an important part of human causal reasoning, and our model easily captures them. Yet a feedforward model, because links only go from input to output, is completely unable to learn such asymmetries and thus is unable to reason asymmetrically.

Captures the Distinction between Necessary and Sufficient Causes

Our ability to model asymmetries in causal learning and reasoning also allows us to capture what has been identified as a central distinction in causal reasoning, the difference between necessary and sufficient causes. For instance, a lit match is sufficient to set gasoline on fire, but it is not necessary because there are other ways in which the gasoline can be ignited. This can be captured in our network by assuming that the strength of a link from cause to effect captures the sufficiency of a cause; the stronger this link the more likely the cause is to bring the effect about. In contrast, the link from effect to cause captures the necessity of a cause; the stronger the link, the more likely it is that the cause preceded the effect. A very strong link from effect to cause suggests that the effect is almost always preceded by that cause, suggesting that the cause is necessary for the effect to come about. Because it cannot learn such asymmetries, having links that only run from cause to effect, a feedforward model cannot learn or use information about this fundamental distinction between necessary and sufficient causes.

Cue competition among effects

In the human and animal causal learning literature, there is considerable evidence for cue competition among causes. One example of such cue competition is Blocking, where first learning that cue A strongly predicts an effect prevents the later learning of the relation between cue B and the effect, even if cue B is highly predictive of the effect. The standard explanation is that cues compete for predictive strength and that when cue A is learned to strongly predict the effect, this essentially captures all the available predictive strength, leaving none for B. Both the Rescorla-Wagner rule and feedforward networks with delta rule learning can capture such cue competition for causes.

But does cue competition for effects also occur, when a single cause predicts multiple effects? Waldmann (1996) points out that the Rescorla-Wagner model strongly predicts such effects and that their absence would create serious problems for this model. However, Waldmann argues that his causal-model theory predicts that cue competition for effects should not occur. And across several studies, he found no evidence for cue competition for effects.

However, other researchers (e.g., Chapman, 1991; Shanks, 1991; Shanks, Lopez, Darby, & Dickinson, 1996) have provided evidence for cue competition for effects. Miller and Matute (1996) have argued that the discrepancy in results among various researchers might be attributable to differences in the questions used to assess causal strength.

This possibility is particularly clear in terms of our model, which suggests that whether one gets cue competition for effects may depend strongly on the type of question that is asked to assess causal strength. That is, does the question ask subjects to assess the strength from cause to effect or from effect to cause? Waldmann (1996), among others, has characterized this difference as between asking predictive questions and asking diagnostic questions. A predictive question asks subjects to assess the extent to which the cause predicts potential effects. In terms of our model, such a question asks subjects to assess the strength of the link from the causes forward to the effect. In contrast, a diagnostic question asks subjects to assess the extent to which the effect is diagnostic of the cause, that is, to what extent the existence of the effect provides evidence for the cause. In terms of our model, this question asks subjects to assess the strength of the link from the effect to the cause.

Cue competition for causes is typically demonstrated when two or more causes predict a single effect, and researchers ask a predictive question about the extent to which the causes predict the effects. Our model suggests that cue competition for effects should be demonstrated when a single cause predicts two or more effects, and subjects are asked a diagnostic question, for which they must assess the strength of the link from the effect back to the cause.

One obvious implication of this is that the learner must be able to separately encode the link from cause to effect, and the link from the effect back to the cause. Several researchers (e.g., Shanks, Lopez, Darby, & Dickinson, 1996) have suggested that this can be captured by assuming two feedforward networks, one that learns the relations from causes to effects and the other which learns the relations from effects back to cause. However, such a solution seems inelegant. With the current model, such an assumption is unnecessary, as a basic part of its architecture is that it can learn separate weights for the two links from cause to effect, and from effect to cause.

Table 2: Learning History for Simulation of Cue Competition

Simulation	Unit	Learning history		Epochs
Phase I	A	+ + + + +		10
	X	. . .		
	Y	+ + + + +		
Phase II	A		+ + + + +	
	X		+ + + + +	10
	Y		+ + + + +	

We have successfully simulated cue competition for effects when a single cause predicts multiple effects and the right question is asked. In the simulation, there are two alternative stimulus presentations (See Table 2). In the first, a single cause (A) is presented that predicts two effects (X and Y) (Phase II alone). In the second, the network is first presented with a number of instances of one effect (Y) predicted by a single cause (A) (Phase I), followed by two effects (X and Y) predicted by the same cause (A) (Phase II).

In this model, separate links are learned from cause to effect, and from effect to cause. And as can be seen in Figure 1, there is an asymmetry in the learned links for the learning sequence of Phase I followed by Phase II. Moreover, it is clear from the links that in this model whether one should expect to get cue competition for effects, depends upon the direction of reasoning. For Phase II alone, equal weights are learned among all the causes and effects (.76). However, when Phase I is presented first, followed by Phase II, the results are quite different, predicting a cue competition effect for effects or consequences. First, there are strong weights from the cause A to both effects X and Y, although the weight is twice as strong from A to Y (1.53 vs. .76). However, in the reverse direction, the weight from X to A is 0, while the weight from Y to A is 1.53. And when we examine the resulting activations (See Table 3) when each of the causes and effects are tested, we get strong evidence for cue competition in backward reasoning from effects to causes, but not in forward reasoning from causes to effects. When effect X is turned on, neither cause A nor effect Y is activated at all. In contrast, when effect Y is turned on, both cause A and effect X are activated. Further, when cause A is activated, effects X and Y have almost identical activations, although Y is slightly higher. Thus, there is strong evidence for cue competition when reasoning backward, from effect to cause, but not when reasoning forward, from cause to effect. Thus, this model suggests that whether one gets cue competition for effects will depend on the direction of reasoning.

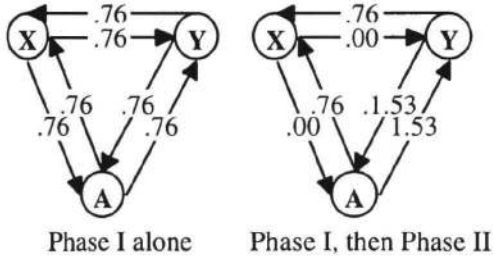

Figure 1: Weights for Cue Competition Simulation

Table 3: Output Activations for Cue Competition Simulation

Simulation	Units Tested	Resulting activations for: A	X	Y
Phase II alone	A	.55	.29	.29
	X	.29	.55	.29
	Y	.29	.29	.55
Phase I followed by Phase II	A	.58	.32	.39
	X	.00	.48	.00
	Y	.39	.32	.58

Represents Learning of and Reasoning with Relations between Causes.

As we noted above, Rescorla-Wagner and feedforward models cannot directly capture relations between causes, but only relations between cause and effect. As a result, we argue that feedforward models are unable to capture Discounting and Augmenting (Jones & Davis, 1965; Kelley, 1971), although feedback models can.

A number of authors (e.g., Baker, Mercier, Vallée-Tourangeau, Frank, & Pan, 1993; Shanks, 1985, 1991; Vallée-Tourangeau, Baker, & Mercier, 1994; Van Overwalle & Van Rooy, 1998) have suggested that a feedforward model with delta rule learning can handle the Discounting and Augmenting principles identified by Kelley (1971) and Jones and Davis (1965). The claim is that Discounting is the same as Blocking found in studies of animal learning, and Augmenting is the same as Super-conditioning. However, despite their apparent similarity the underlying processing mechanisms for the two sets of phenomena are quite different. Moreover, feedforward models lack the necessary mechanism for capturing discounting and augmenting, as they lack the ability to represent relations among causes, which we argue is critical for capturing these effects..

In Blocking, if the organism first learns that A is strongly associated with effect X, when it is later presented examples of B and A covarying with X, the organism fails to learn the new association between B and X. In terms of error correcting learning, such as the delta-rule, once X is strongly predicted by A, when A and B are subsequently paired with X, there is little discrepancy between the actual and predicted value of X (no error) and therefore little change is made in the weight from B to X.

Thus, Blocking clearly deals with competition in the initial learning of the causal links. In contrast, Discounting in the human literature clearly deals with competition among already learned causal explanations. Kelley (1971) and Jones and Davis (1965) were considering adults who were relying on already learned and activated knowledge. For instance, consider adults who are told that a woman wrote a pro-abortion essay after being assigned to the position by her debate coach. Because of the assignment, they should discount a pro-abortion attitude as a cause of her behavior. These adults already know that both a pro-abortion attitude and the assignment by the coach are possible explanations for the behavior. They are not learning these relationships for the first time. Thus, in contrast to Blocking, Discounting does not refer to the failure to learn a causal link, but rather reasoning on the basis of already learned causal knowledge.

What changes in the typical Discounting situation is information about the availability or presence of a potential cause in a particular situation. Both McClure (1998) and Morris and Larrick (1996) have argued that the degree of discounting between two causes is a function of the extent to which they are positively or negatively related. Discounting can be handled in an auto-associative model by assuming that there is an inhibitory link between competing explanations (Read & Miller, 1993; Read & Marcus-Newhall, 1993) (This cannot be done in a feedforward model). Because of the inhibitory link, increased availability of a plausible alternative will reduce the activation of the other explanation. Thus, we aren't looking at competition for learning of links, but rather at competition for the activation of concepts with previously learned causal links.

Now consider Super conditioning and its relation to Augmenting. If the organism learns that A is followed by X, but A and B together are not followed by X, then B develops a negative or inhibitory relationship with X. If the organism then learns that D and B together are followed by X, then the relationship between D and X becomes stronger than it would have been if B had not first developed a negative relationship with X. Again, although this phenomena is similar to Augmenting, it is not the same thing. Augmenting deals with inhibition between an already learned cause and effect, whereas Super-conditioning is based on inhibition in the initial learning of causal relationships. For example, suppose we are told that someone got an A on an extremely difficult exam. We use our preexisting causal knowledge to infer that the individual must be quite smart. We are clearly not learning for the first time that someone who can overcome a major barrier must possess a considerable amount of the relevant ability, which is what Super-conditioning would be concerned with. Clearly, there is a critical distinction between the initial acquisition of information and the ways in which it is later used.

Morris and Larrick (1996) make a similar distinction. They note that in models of causal reasoning, there is a distinction between induction or the initial acquisition of causal knowledge, and reasoning or attribution, the actual use of that knowledge. For instance, Kelley's (1971) ANOVA cube model is a model of the acquisition of causal knowledge, whereas his causal schema model is a model of the use of pre-existing knowledge for reasoning.

Thus, the two types of phenomena are fundamentally different in terms of the underlying processing mechanisms. Blocking and Super-conditioning deal with competition for weight strength in the learning of new causal relations, whereas Discounting and Augmenting deal with competition for activation in the use of already learned causal relations. These are quite different processes. And as we noted, a feedforward model is unable to capture a situation in which the causal mechanism depends on links among causes.

Summary

In this paper we have demonstrated that a feedback or recurrent, auto-associative model can capture several important aspects of causal learning and causal reasoning that cannot be handled by the feedforward models that have been the typical focus of investigation. First, our model can learn asymmetric relations between cause and effect. Second, it can reason in both directions between cause and effect. As a result it can represent an important distinction in causal reasoning, the difference between necessary and sufficient causes. Third, because the nodes in the network are totally interconnected, it can represent cue competition among effects, something which can only be done with feedforward models by assuming that two separate networks are learned, a highly non parsimonious assumption. Finally, we argue that contrary to previous claims, a feedforward model cannot handle Discounting and Augmenting in causal reasoning. However, a feedforward model can. The success of our feedback model suggests that researchers should focus more energy on the capabilities of such models of causal learning and reasoning.

References

Baker, A. G., Mercier, P., Vallée-Tourangeau, F., Frank, R., & Pan, M. (1993). Selective associations and causality judgments: Presence of a strong causal factor may reduce judgments of a weaker one. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19, 414-432.

Chapman, G. B. (1991). Trial order affects cue interaction in contingency judgment. Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 837-854.

Chapman, G. B., & Robbins, S. J. (1990). Cue interaction in human contingency judgment. Memory and Cognition, 18, 537-545.

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. Journal of Personality and Social Psychology, 58, 545-567.

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

Gluck, M. A., & Bower, G. H. (1988). Evaluating an adaptive network model of human learning. Journal of Memory and Language, 27, 166-195.

Jones, E. E. (1990). Interpersonal perception. New York: W. H. Freeman.

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). New York: Academic Press.

Kelley, H. H. (1971). Attribution in social interaction. In E. E. Jones, D. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), Attribution: Perceiving the causes of behavior. Morristown, NJ: General Learning Press.

McClelland, J. L., & Rumelhart, D. E. (1986). (Eds.). Parallel Distributed Processing: Explorations in the microstructure of cognition. Vol. 2: Psychological and Biological Models. Cambridge, MA: MIT Press/Bradford Books.

McClelland, J. L., & Rumelhart, D. E. (1988). Explorations in parallel distributed processing: A handbook of models, programs, and exercises. Cambridge, MA: MIT Press/Bradford Books.

McClure, J. (1998). Discounting causes of behavior: Are two reasons better than one? Journal of Personality and Social Psychology, 74, 7-20.

Miller, R. R., & Matute, H. (1996). Animal analogues of causal judgment. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), The Psychology of Learning and Motivation, Vol. 34: Causal learning. San Diego, CA: Academic Press.

Montoya, J. A., & Read, S. J. (1998). A constraint satisfaction model of the correspondence bias: The role of accessibility and applicability of explanations. In M. A. Gernsbacher & S. J. Derry (Eds.), Proceedings of the Twentieth Annual Conference of the Cognitive Science Society. Mahwah, NJ; Erlbaum.

Morris, M. W., & Larrick, R. P. (1996). When one cause casts doubt on another: A normative analysis of discounting in causal attribution. Psychological Review,

102, 331-355.

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. Ablex: Norwood, NJ.

Read, S. J., Lincer-Hill, H.(1999). Principles of explanatory coherence in trait inferences. Unpublished manuscript, University of Southern California, Los Angeles, CA.

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, 526-540.

Read, S. J., & Miller, L. C. (1994). Dissonance and balance in belief systems: The promise of parallel constraint satisfaction processes and connectionist modeling approaches. In R. C. Schank & E. Langer (Eds.), Beliefs, reasoning, and decision making: Psycho-logic in honor of Bob Abelson). Hillsdale, NJ: Lawrence Erlbaum Associates.

Read, S. J., & Miller, L. C. (1998). On the dynamic construction of meaning: An interactive activation and competition model of social perception. In S. J. Read & L. C. Miller (Eds.) Connectionist models of social reasoning and behavior.(pp. 27-68). Mahwah, NJ: Erlbaum.

Read, S. J., & Montoya, J. A. (in press). An autoassociative model of causal reasoning and causal learning: Response to Van Overwalle's critique of Read and Marcus-Newhall (1993). Journal of Personality and Social Psychology.

Read, S. J., Vanman, E. J., & Miller, L. C. (1997). Connectionism, parallel constraint satisfaction processes, and gestalt principles: (Re)introducing cognitive dynamics to social psychology. Personality and Social Psychology Review, 1(1), 26-53.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), Classical conditioning II: Current research and theory. New York: Appleton-Century-Crofts.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), Advances in experimental social psychology. (Vol. 10). New York: Academic Press.

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, P.K., & Ranney, M. (1991). An empirical investigation of the psychological fidelity of ECHO: Modeling and experimental study of explanatory coherence. Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum.

Schank, P. K., & 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. . Hillsdale, NJ: Erlbaum.

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgment. Quarterly Journal of Experimental Psychology, 37B, 1-21.

Shanks, D. R. (1991). Categorization by a connectionist network. Journal of Experimental Psychology: Learning, Memory, and Cognition. 17, 433-443.

Shanks, D. R., Lopez, F. J., Darby, R. J., & Dickinson, A. (1996). Distinguishing associative and probabilistic contrast theories of human contingency judgment. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), The Psychology of Learning and Motivation, Vol. 34: Causal learning. San Diego, CA: Academic Press.

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

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

Vallée-Tourangeau, F., Baker, A. G., & Mercier, P. (1994). Discounting in causality and covariation judgments. Quarterly Journal of Experimental Psychology, 47B, 151-171.

Van Overwalle, F. (1998). Causal explanation as constraint satisfaction: A critique and a feedforward connectionist alternative. Journal of Personality and Social Psychology, 74, 312-328.

Van Overwalle, F., & Van Rooy, D. (1998). A connectionist approach to causal attribution. In S. J. Read & L. C. Miller (Eds.) Connectionist models of social reasoning and behavior. Mahwah, NJ: Erlbaum.

Waldmann, (1996). Knowledge-based causal induction. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), The Psychology of Learning and Motivation, Vol. 34: Causal learning. San Diego, CA: Academic Press.

Wang, H., Johnson, T. R., & Zhang, J. (1997). UEcho: A model of uncertainty management in human abductive reasoning. In M. G. Shafto & P. Langley (Eds.). Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society. Mahwah, NJ: Erlbaum.

Widrow, G., & Hoff, M. E. (1960). Adaptive switching circuits. Institute of Radio Engineers, Western Electronic Show and Convention, Convention Record, Part 4, 96-104