Psychological Review
1996, Vol. 103, No. 2, 219-240

0033-295X/96/$3.00

Cognitive Dissonance Reduction as Constraint Satisfaction

Thomas R. Shultz
McGill University

Mark R. Lepper
Stanford University

A constraint satisfaction neural network model (the consonance model) simulated data from the two major cognitive dissonance paradigms of insufficient justification and free choice. In several cases, the model fit the human data better than did cognitive dissonance theory. Superior fits were due to the inclusion of constraints that were not part of dissonance theory and to the increased precision inherent to this computational approach. Predictions generated by the model for a free choice between undesirable alternatives were confirmed in a new psychological experiment. The success of the consonance model underscores important, unforeseen similarities between what had been formerly regarded as the rather exotic process of dissonance reduction and a variety of other, more mundane psychological processes. Many of these processes can be understood as the progressive application of constraints supplied by beliefs and attitudes.

A foolish consistency is the hobgoblin of little minds.
—Ralph Waldo Emerson, *Essays: First Series*

Consistency is the last refuge of the unimaginative.
—Oscar Wilde, "The Relation of Dress to Art"

Despite its apparent poor reputation in literary circles, as an indication of intellectual mediocrity, cognitive consistency has a long and distinguished history of study in psychology. Taken as a general sign of human rationality, seeking consistency among beliefs and attitudes has held a prominent place in social-psychological theorizing for half a century (e.g., Festinger, 1957; Heider, 1946, 1958; McGuire, 1960; Newcomb, 1953; Osgood & Tannenbaum, 1955). This widespread concern with issues of cognitive consistency culminated in the late 1960s in the publication of the 84-chapter *Theories of Cognitive Consistency,* edited by six of the major theorists of the day (Abelson, Aronson, McGuire, Newcomb, Rosenberg, & Tannenbaum, 1968). In recent years, however, the study of cognitive consistency seems to have fallen out of favor, perhaps in part because of an inability to further penetrate its underlying reasoning mechanisms.

In this article we re-visit the classic problem of cognitive consistency, drawing on more recent advances in the use of parallel processing techniques to model psychological phenomena. We argue that the reduction of cognitive dissonance, and more generally the search for cognitive consistency, can be usefully viewed as a constraint satisfaction problem, and we propose a computational model of the process of consonance seeking. Simulations using this consonance model are then shown to capture data from a number of classic and prototypic dissonance experiments, often more effectively than dissonance theory itself, and predictions from this model are tested against data from a new dissonance experiment undertaken for this purpose.

Cognitive Dissonance

Cognitive dissonance theory (Festinger, 1957) has been a pillar of social psychology for nearly 40 years. The theory assumes that dissonance is a psychological state of tension that people are motivated to reduce. Any two cognitions are dissonant when, considered by themselves, one of them follows from the obverse of the other. The amount of dissonance is a function of the ratio of dissonant to total relevant (i.e., dissonant plus consonant) relations, with each relation weighted for its importance to the person. Dissonance can be reduced by decreasing the number or the importance of the dissonant relations, or both, or by increasing the number or the importance of consonant relations. How dissonance actually gets reduced depends on the resistance to change of the various relevant cognitions, with less resistant cognitions being more likely to change. Resistance derives from the extent to which a change would produce new dissonance, the degree to which the cognition is anchored in reality, and the difficulty of changing those aspects of reality.

Festinger (1957) used dissonance theory to account for a wide array of psychological phenomena, including the transmission of rumors, rationalization of decisions, attitudinal consequences of counterattitudinal advocacy, selectivity in information search and interpretation, and responses to the disconfirmation of beliefs. It has since been successfully applied to many phenomena in a wide variety of both predictive and postdictive contexts (e.g., Aronson, 1969; Brehm & Cohen, 1962; Cooper & Fazio, 1984; Festinger, 1964; Greenwald & Ronis, 1978; Lepper, 1983; Steele, 1988; Wicklund & Brehm, 1976; Zimbardo, 1969).

This research was supported by grants from the Social Sciences and Humanities Research Council of Canada and the U.S. National Institute of Mental Health.

Denis Mareschal, Stephen Read, Barbara Spellman, and Sheldon Tetewsky commented helpfully on earlier drafts.

Correspondence concerning this article should be addressed to Thomas R. Shultz, Department of Psychology, McGill University, 1205 Penfield Avenue, Montréal, Québec, Canada H3A 1B1, or to Mark R. Lepper, Department of Psychology, Stanford University, Jordan Hall, Building 420, Stanford, California 94305-2130. Electronic mail may be sent to shultz@psych.mcgill.ca or to lepper@psych.stanford.edu.

Consonance Model

In this article we present a computational model of cognitive dissonance that we refer to as the *consonance model.* There are several potential payoffs for developing this sort of model. Such a model would not only provide a novel explanation for the many existing dissonance phenomena but might also lead to the prediction of new dissonance phenomena. Such a model would also promote the theoretical unification of cognitive dissonance findings with other psychological phenomena by providing a common conceptual framework for examining the historically distinctive and seemingly exotic process of dissonance reduction and a variety of more mundane cognitive processes.

Psychological Justification

The model is based on the idea that dissonance reduction can be viewed as a constraint satisfaction problem. In other words, the motive to seek cognitive consistency postulated by dissonance theory and related models (e.g., Abelson et al., 1968; Feldman, 1966) can be seen as imposing constraints on the beliefs and attitudes that an individual holds simultaneously (cf. Abelson & Rosenberg, 1958). Such problems can be solved by the simultaneous satisfaction of many soft constraints that can vary in their relative importance. Soft, as opposed to hard, constraints are those that are desirable, but not essential, to satisfy. The striving for complete cognitive consistency, the theoretical but never-attained goal of dissonance reduction, might well involve constraints of this sort.

Within this framework, the networks used in our model correspond to a subject's representation of the situation created, or the psychological problem posed, by the experimental settings in the classic cognitive dissonance paradigms. In these networks, activations of various units represent the direction and strength of the individual's beliefs and attitudes (including beliefs and attitudes regarding the person's own actions). Units may also differ in their resistance to change, reflecting differences in the extent to which cognitions may be supported by other cognitions or anchored in reality. Connection weights between cognitions represent psychological implications among the person's beliefs and attitudes. The connections between any two units, as between any two cognitions in the classic dissonance model, can be either positive or negative, or the two may be psychologically irrelevant to one another. Both unit activations and weights may vary, depending on the paradigm, across the different conditions of a single experiment. Increasing consonance—conceptually, the degree to which similarly evaluated units are linked by positive weights and oppositely valued units are linked by negative weights—corresponds to the process of reducing dissonance, or striving for consistency, among personal beliefs and attitudes.

Computational Instantiation

Constraint satisfaction networks have been shown to be capable of simulating a variety of phenomena in cognitive psychology, including belief revision, explanation, schema completion, analogical reasoning, causal attribution, discourse comprehension, and content-addressable memories (Holyoak & Thagard, 1989; Kintsch, 1988; Rumelhart, Smolensky, McClelland, & Hinton, 1986; Shultz, 1992; Sloman, 1990; Thagard, 1989). Constraint satisfaction networks are also beginning to be applied to a variety of phenomena in social psychology, including attitude change, cognitive balance, and cognitive dissonance (Read & Miller, 1994; Shultz & Lepper, 1992; Spellman & Holyoak, 1992; Spellman, Ullman, & Holyoak, 1993). Unless used to model long-term memory, these networks are generally considered ephemeral, in the sense that they are created on-line to deal with some particular task. The process of creating the network is not usually modeled, presumably because it is not sufficiently understood psychologically. These networks function by reducing energy (or equivalently, maximizing goodness) subject to the constraints supplied by the connections and any external input. Our consonance model for reducing cognitive dissonance is a constraint satisfaction network lacking some of the parameters of other such networks and introducing some special parameters of its own.

Hopfield (1982, 1984) worked out the mathematics for solving constraint satisfaction problems in parallel networks. Maximizing the consonance (or goodness) of any pair of connected units depends on the sign of the connection between them. For purposes of illustration, assume an activation range for units of 0 to 1, the range actually used in our simulations. If connected by a positive weight, both units of the pair should be active to maximize consonance. With a negative weight, consonance is maximized when the two units are not both active, that is, when both are inactive or only one is active. Activations will change over time cycles so as to satisfy the various constraints and increase consonance.

More formally, the consonance contributed by a particular unit i is as follows:

$$\text{consonance}_i = \sum_j w_{ij} a_i a_j, \tag{1}$$

where w_{ij} is the weight between units i and j, a_i is the activation of the receiving unit i, and a_j is the activation of the sending unit j.[1]

The overall consonance in the network is the sum of the values given by Equation 1 over all receiving units in the network:

$$\text{consonance}_n = \sum_i \sum_j w_{ij} a_i a_j. \tag{2}$$

Activation spreads over time cycles by two update rules:

$$a_i(t+1) = a_i(t) + \text{net}_i(\text{ceiling} - a_i(t)), \text{ when } \text{net}_i \geq 0, \tag{3}$$

and

$$a_i(t+1) = a_i(t) + \text{net}_i(a_i(t) - \text{floor}), \text{ when } \text{net}_i < 0, \tag{4}$$

where $a_i(t+1)$ is the activation of unit i at time $t+1$, $a_i(t)$ is the activation of unit i at time t, ceiling is the maximal level of activation, floor is the minimal activation, and net_i is the net input to unit i, defined as

[1] For now, dissonance can be considered as negative consonance. Later, in Equation 6, we formalize dissonance as negative consonance, standardized for number of intercognition relations.

$$\text{net}_i = \text{resist}_i \sum_j w_{ij} a_j. \quad (5)$$

The parameter resist_i is a measure of the resistance of receiving unit i to having its activation changed. The larger the value of the resistance multiplier, the less the resistance to change.[2]

At each time cycle, n units are randomly selected and updated according to Equations 3 and 4. By default, n is the number of units in the network. This updating scheme allows a unit to be updated more or less than once per cycle, but on the average each unit will be updated about once per cycle. Random selection of units to update increases variability across networks, which might correspond to individual differences among human subjects. The update rules ensure that consonance increases or at least stays the same across cycles. At some point, consonance reaches a local maximum asymptotic value, from which it can no longer increase. At this point, the updating process is stopped. A few additional parameters concerning the construction of the networks are described in the next section.

Mapping Dissonance Theory to the Consonance Model

Having described both dissonance theory and the consonance constraint satisfaction model, we now provide a systematic mapping between the two. This mapping, in turn, specifies how the various consonance simulations were conducted. The mapping process is organized around a series of five theoretical principles. To varying degrees, these theoretical principles were specified in classical dissonance theory. Additional specifications, where necessary, are supplied by the consonance model. Each theoretical principle governs the design of all simulations with the consonance model.

This mapping exercise is similar in spirit to those provided in previous constraint satisfaction models of analogical mapping, explanatory coherence, and attitude change. Holyoak and Thagard (1989) translated structural, semantic, and pragmatic principles into constraint satisfaction networks that mapped source-to-target analogies. Thagard (1989) mapped seven principles for coherence among propositions to a constraint satisfaction model of hypothesis evaluation. Spellman and Holyoak (1992; Spellman et al., 1993) applied similar principles to study the structure and change of American attitudes toward the Persian Gulf war. Although there are fundamental similarities in the mathematics of constraint satisfaction across these different simulation projects, the mapping of domain-specific theoretical ideas to constraint satisfaction principles does vary significantly with the particular domain being mapped.

Mapping Relation 1: Cognitions

The basic units in dissonance theory are cognitions. Cognitions are not fully specified in dissonance theory but appear to include both beliefs and evaluations (i.e., attitudes). Such cognitions could be assumed (beyond dissonance theory) to vary in both direction and strength. The positive direction could represent that something is either believed to be true or is favorably evaluated. Analogously, the negative direction could represent that something is either believed to be false or is negatively evaluated. Strength is the degree to which something is believed to be true or false or evaluated positively or negatively. For example, one might highly value one's spouse but only weakly believe that the Chicago Cubs will win baseball's World Series.[3]

Manipulations within dissonance experiments are designed to ensure that subjects begin the experiment with certain beliefs and evaluations. For example, in the so-called forbidden toy studies, children are given either a mild or a severe threat against playing with a particularly desirable toy (Aronson & Carlsmith, 1963; Freedman, 1965). Thus, it can be assumed that these children begin the experiment with a positive evaluation of the toy and a belief that they were just given a particular level of threat not to play with the toy.

In the consonance model, a cognition is represented by the net activation of a pair of negatively connected units. One unit represents the positive direction and the other represents the negative direction. Net activation for the cognition equals the difference between activation of the positive unit and activation of the negative unit. Anderson (1995, pp. 150–152) has reviewed neurological and computational rationales for using pairs of units in this way. Briefly, neurons are sometimes organized into excitatory and inhibitory camps that respond in opposite ways to input, one group being excited and the other group being inhibited by the same input. Furthermore, the activation range for positive neurons is typically greater for positive than for negative neurons. Mimicking these principles, our *ceiling* activation parameter is set to 1 for units representing positive aspects of cognitions, and to 0.5 for units representing negative aspects of cognitions; we refer to the lower ceiling on negative units as the *minus ceiling* parameter. Our *floor* activation parameter is set to 0 for both types of units.

This representational scheme allows for some degree of ambivalence in cognitions. For example, there might be some evidence favoring a belief and other evidence against the belief; or something might be both liked and disliked. Recent research on attitude measurement has stressed the importance of ambivalence (e.g., Thompson, Zanna, & Griffin, 1995). The negative weight between the paired units tends to discourage such ambivalence as activation on one end of the dimension increases and then drives down activation on the other end (following the update rules in Equations 3, 4, and 5). However, relatively persistent ambivalence could be produced if both the positive and negative units for a cognition receive strong support from other cognitions. Such ambivalence creates dissonance as explained in the following section, *Mapping Relation 2: Elementary Dissonance.*

To simulate particular dissonance experiments, we provide initial activation values for particular cognitions reflecting the various experimental manipulations. Generally, these initial activations have default values of 0.5 for high and 0.1 for low. If there is a strong belief that a behavior was engaged in or that

[2] Our resistance parameter is mathematically identical to the *istr* parameter used by Rumelhart et al. (1986) to scale the importance of internal network contributions to activation updates, although these two parameters are given quite distinct psychological interpretations.

[3] Truth and evaluation are different and perhaps need to be treated differently in some contexts. In the present model, it is not necessary to distinguish them. In this context, it is interesting to note that early cognitive consistency theorists (Festinger, 1957; Heider, 1958) also treated truth and evaluation in a similar fashion.

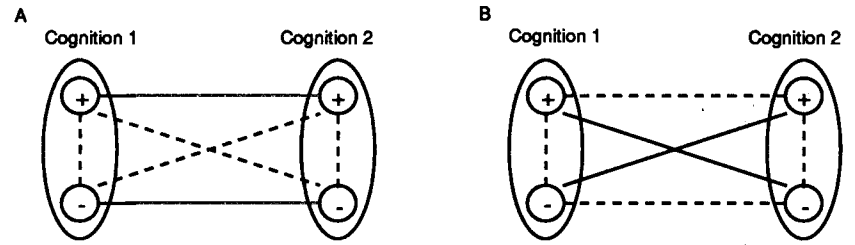

Figure 1. Any two cognitions can be connected positively (as shown in Figure 1A), negatively (as shown in Figure 1B), or can be unrelated. In this figure, positive connection weights are symbolized by solid lines, negative connection weights by dashed lines. Each cognition is symbolized by an ellipse drawn around the positive and negative poles of the cognition. See text for further explanation.

something is highly valued, the positive unit of the corresponding cognition would have an initial activation of 0.5. For example, simulations of the forbidden toy studies begin with evaluation of the desirable toy at a net activation of 0.5. If there is a strong belief that a behavior was not engaged in or that something is highly disliked, the negative unit of the corresponding cognition would have an initial activation of 0.5. For example, simulations of the forbidden toy studies begin with the *play with toy* cognition at a net activation of −0.5 to reflect that no child ever played with that toy (0 − 0.5 = −0.5).

If an experimental manipulation is provided in two different amounts (e.g., high and low), this difference is reflected in differential initial activations for the conditions. For example, in simulations of the forbidden toy studies, the cognition that one has been threatened is given an initial net activation of 0.5 (high) in the severe threat condition and 0.1 (low) in the mild threat condition.

Mapping Relation 2: Elementary Dissonance

In dissonance theory, two cognitions taken by themselves are said to be dissonant when one follows from the obverse of the other. Conversely, two cognitions are consonant when one implies the other. Although the nature of these implicational relations is not fully specified in dissonance theory, it appears broad enough to include logical implication, causal relations, psychological implication, expectation, and association (cf. Abelson, 1968; Aronson, 1969).

In the consonance model, cognitions (represented as net activations in pairs of negatively connected units) are connected to other such cognitions to form a network representing a person's relevant beliefs and attitudes regarding a particular experimental situation. A negative implication is represented by inhibitory weights between two cognitions; a positive implication is represented by excitatory weights between two cognitions. Connection weights can range from −1 to +1. The absolute default value for high weights is 0.5; that for low weights is 0.1 (although we do not use low weights in any simulation reported here).

Any two cognitions can be positively related, negatively related, or unrelated. Unrelated cognitions have connection weights of 0. The positive and negative connection schemes for two generic cognitions are illustrated in Figures 1A and 1B, respectively. For two cognitions that are positively related (Figure 1A), their positive poles are linked with excitatory weights, as are their negative poles; inhibitory weights link the positive pole of one cognition with the negative pole of the other cognition. Such connections are reversed for cognitions that are negatively related, as shown in Figure 1B. In both cases, the positive and negative poles of each cognition are connected with inhibitory weights, each unit has an inhibitory self-connection specified by the *cap* parameter (not shown in Figure 1), and all connection weights are bidirectional (also not shown). Again, connection weights have a default value of 0.5.

The consonance contributed by a particular unit in a network is given by Equation 1, the sum of the triple products of the activation on the unit times the activation on each sending unit times the corresponding weight between the sending and receiving units. Dissonance can be considered as the negative of consonance.

Equation 1 ensures that consonant relations are produced by cognitive patterns yielding positive triple products, whereas dissonant relations are produced by cognitive patterns yielding negative triple products. As shown in Table 1, positive triple products (consonant relations) are the result of either one or three positive directional signs; negative triple products (dissonant relations) are the result of either one or three negative directional signs.

Some readers may find the relations in Table 1 reminiscent of cognitive balance theory (Heider, 1958), especially as elaborated by Abelson and Rosenberg (1958; Rosenberg & Abelson, 1960) and by Cartwright and Harary (1956) to apply to larger networks of cognitions. Our analysis is in the spirit of these earlier attempts to provide a more formal and general mathemati-

Table 1
Qualitative Relations Generated by Equation 1

Sending unit activation	Connection weight	Receiving unit activation	Relation
+	+	+	Consonant
−	+	−	Consonant
+	−	−	Consonant
−	−	+	Consonant
−	−	−	Dissonant
+	−	+	Dissonant
−	+	+	Dissonant
+	+	−	Dissonant

cal representation of cognitive consistency, although our analysis goes beyond these earlier efforts in several respects. First, Equation 1 generates quantitative values of which the relations in balance theory are qualitative idealizations. That is, Equation 1 indicates how consonant or how dissonant a particular relation is. In a similar way, our consonance model specifies quantitatively the amount of change predicted for specific cognitive elements. Finally, because each cognition in our consonance model is represented by a negatively connected pair of units, Equation 1 also deals with subcognition relations, within a particular cognition or between parts of two cognitions. Notice that our two-unit per cognition representational scheme allows for dissonance (or consonance) within a particular cognition. For example, if there are both positive and negative feelings for the same object, this would be computed as dissonant by Equation 1 because of the inhibitory weight linking the positive and negative units representing this cognition.

The design of connecting weights in our consonance network is accomplished by reading the descriptions of implications provided in the dissonance experiments to be simulated. We always give the implication a high connection weight (default value of 0.5) and always focus on the implication specified in the cause-to-effect direction. For example, in simulating the forbidden toy studies (Aronson & Carlsmith, 1963; Freedman, 1965), there is an implication that a high evaluation of the toy causes one to want to play with the toy. This leads us to link the *toy evaluation* and *play with toy* cognitions with an excitatory relation. Considering the reverse, effect to cause, direction is often more problematic for determining the sign of the relation. For example, wanting to play with a toy has no particular causal implication for evaluation of the toy, making it more difficult to determine appropriate relation linking those two cognitions. One might want to play with a toy out of sheer curiosity about how good the toy is, or out of boredom or a lack of alternatives. Nonetheless, following the usual practice in constraint satisfaction networks, we assign the same default value for relations going in both directions. The point here is simply that the decision about the sign of these relations is dictated by the expected causal direction. It is interesting that other independently derived constraint satisfaction models of cognitive consistency also focus on causal implications in designing networks (Read & Miller, 1994). This focus on causal relations in dissonance models reflects the central importance of causal relations in the understanding of social action (Schank & Abelson, 1977).

Mapping Relation 3: Total Dissonance

Dissonance theory holds that the total amount of dissonance is a function of the ratio of dissonant cognitions to all relevant cognitions (dissonant plus consonant cognitions), with cognitions and relations weighted for their importance to the person.

In the consonance model, total consonance is represented by summing the values computed by Equation 1 over all units in the network. This is specified in Equation 2. Notice that the weighting of relations (weights) and cognitions (activations) is represented in Equations 1 and 2 by the triple products of sending activation, receiving activation, and connecting weight that are summed. The more important any of these things are in terms of believability or value, the larger their numeric value, and the larger their impact on consonance. Notice, too, that irrelevant cognitions (those connected with weights of 0) contribute nothing to consonance. Total dissonance is formally defined as the negative of total consonance divided by r, the number of nonzero intercognition relations in the network.

$$\text{dissonance} = \frac{-\sum_i \sum_j w_{ij} a_i a_j}{r} \tag{6}$$

Dividing by r standardizes dissonance to facilitate comparisons across networks, by controlling for the number of nonzero relations. Otherwise, larger networks or networks with more nonzero relations have the potential of generating more consonance (or more dissonance), merely by virtue of their greater size or greater number of relations. Self-connections, identified by w_{ii}, are excluded from this computation of dissonance so that dissonance does not vary directly with degree of activation.

This definition of dissonance is analogous to energy in constraint satisfaction networks that function by minimizing energy rather than maximizing goodness (e.g., Hopfield, 1982, 1984). Dissonance is to consonance as energy is to goodness. Our definition of dissonance goes beyond the ratio definition of Festinger (1957) because it is formalized, assesses the amount of dissonance in each intercognition relation, includes within-cognition ambivalence, includes "consonant" relations with positive triple products, and can vary even when all relations are dissonant (i.e., all triple products are negative) or all relations are consonant (i.e., all triple products are positive).

Mapping Relation 4: Motivation to Reduce Dissonance

Dissonance theory specifies that people are strongly motivated to reduce cognitive dissonance. This does not imply that people always succeed in doing so, only that there is motivational pressure in the direction of dissonance reduction. Other factors, such as reality, or the potential for creation of new dissonances often stand in the way of complete dissonance reduction.

In the consonance model, networks tend to settle into more stable, less dissonant states as unit activations are updated according to Equations 3, 4, and 5. The various constraints supplied by weights, initial activations, and resistances are all soft constraints. None of them absolutely must be satisfied, but the update rules attempt to satisfy as many of them as possible, as well as possible. There is no guarantee that consonance reaches a global maximum, only that it increases or stays the same. It could, for example, become stuck in a local maximum. Being stuck in a local maximum would correspond to reducing dissonance only up to a point, rather than completely, or to "satisficing" rather than optimizing.

We do three additional things to increase psychological realism in our consonance networks. We discourage unit activation values from reaching their maximum values, randomize all default parameter values, and introduce connection weight asymmetries.

A *cap* parameter, when set to a high negative proportion, prevents activations from growing to their ceiling. Our default setting for cap is −0.5. Mathematically, cap is the value of the con-

nection between each unit and itself, w_{ii}.[4] Hopfield (1982, 1984), who formalized the mathematics underlying constraint satisfaction networks, had assumed that such self-connections are equal to 0. Allowing self-connections to be other than 0 produces additional spurious states in the neighborhood of a desired attractor,[5] thus increasing the variability of solutions (Hertz, Krogh, & Palmer, 1991); that is, a network could fall into a series of different local consonance maxima. We use cap to enforce the psychologically realistic assumption that the events in most dissonance experiments are not of central importance to the subjects. Therefore, activations, particularly those representing evaluations, should be discouraged from reaching maximal values. This limitation on the extremity of evaluations becomes important in some, but not all, of the simulations reported here.

Robustness against parameter variation is sometimes assessed in connectionist network models by doubling or halving a parameter and repeating the simulation (Schneider, 1988). Because of the large number of parameters and simulations in our project, this technique would generate an unmanageably large simulation space.

Instead, as a network is set up, weights, resistances, caps, and initial activations are all randomized by adding or subtracting a random proportion of their initial amounts. A parameter we call *rand%* specifies the proportion range in which additions or subtractions are randomly selected under a uniform distribution. Specifically, an adjusted parameter value is computed as follows:

$$y = x \pm \{\text{random}(\text{absolute}[x \times \text{rand\%}])\}, \qquad (7)$$

where y is the adjusted parameter value, x is the original parameter value, *random* is a procedure that generates a random uniform distribution between 0 and its argument, and *absolute* returns the absolute value of its argument. Our simulations were run under three values of rand%: small (0.1), medium (0.5), and large (1.0). It is worth stressing that initial settings of all weight, resistance, cap, and initial activation parameters were adjusted by this randomization process in each run of each simulation. This procedure is more comprehensive than the traditional sensitivity tests that focus on a few individual parameters, but it will not indicate the precise impact of each parameter. Instead, the present technique is designed to assess efficiently the robustness of the simulations against general parameter variation.

This random perturbation of parameters also serves to introduce some degree of psychological realism because it can be assumed that not everyone shares precisely the same parameter values. Such variation is not necessary in these simulations, however, to capture dissonance phenomena.

The randomization of weight values violates the weight symmetry assumed by Hopfield (1982, 1984), in that $w_{ij} \neq w_{ji}$. Hopfield reported that violations of the symmetry assumption increased memory errors and instability in network solutions to memory retrieval problems. Such results may also correspond to natural psychological variation.

In summary, randomization of parameter values and use of the cap and rand% parameters discourage units from reaching extreme activation values and increase the variability of network solutions, thus increasing psychological realism.

Mapping Relation 5: Resistance and Modes of Dissonance Reduction

Festinger (1957) specified that dissonance can be reduced by decreasing the number or importance of dissonant relations or by increasing the number or importance of consonant relations, or a combination of these factors. Presumably, dissonance reduction could be accomplished by changing evaluations, beliefs, or implications among them. However, dissonance theory also specified that the cognitions most likely to change are those least resistant to change. Resistance stems from the possible creation of new dissonance because of relations with other cognitions, from cognitions that are anchored in reality, and from the difficulty of changing aspects of reality. In practice, in dissonance experiments, this has typically meant that dissonance is reduced by changes in evaluations, not by changes in beliefs about salient events that have just happened in the experimental setting, nor by changes in implications among cognitions.

In the consonance model, implications among cognitions, represented by connection weights, never change as the network settles. Only unit activations are allowed to change. Change in activations is strongly affected by a resistance parameter, a scalar that modulates the net input to a receiving unit, as specified in Equation 5.

In a more complete model, resistance might well be implemented by constraining connections to many other beliefs. For simplification, we implement this with an explicit resistance parameter. Our default values for high and low resistances throughout the present simulations are 0.01 and 0.5, respectively. Recall that the larger the resistance multiplier, the more readily the unit will change its activation. Thus, larger resistance multipliers implement lower resistances.

In all of the consonance models of dissonance experiments reported here, we give evaluative cognitions low resistance and belief cognitions high resistance. This reflects the fact that participants in the original experiments were undoubtedly quite certain of what had just happened to them, but were probably somewhat unclear about their evaluations of particular novel features of the experiment.

The Generic Consonance Network

All of the simulations presented here use a generic consonance network. The cognitions in dissonance experiments fall into one of three categories: behaviors, justifications, or evaluations. In the classic insufficient justification paradigms, as will be seen, there is one behavior, one justification for the behavior, and one evaluation. In the traditional free-choice paradigm, there is one behavior (a decision to choose one alternative over another) and two evaluations (one referring to each alternative). Our consonance program enables specification of each of the relevant cognitions, including their type and their initial activations, and of the relations among the cognitions. Different dissonance experiments require different instantiations of this generic network because they involve differ-

[4] Thanks to Denis Mareschal for this suggestion.

[5] In the present context, a network attractor corresponds to a particular consonant set of relations among the attitudes and beliefs represented in the network.

ent particular types of cognitions, with differing particular initial activation values, and particular implications among cognitions. As already noted, evaluation cognitions are given low resistance, whereas other cognition types (about behaviors and justifications) are given high resistance.

An Example of Activation Updates

Before proceeding to the simulations, some readers might appreciate a concrete example of how unit activations change over time cycles. Consider a simple example of two positively related cognitions, both of which are evaluations, with low resistance to change. Default high connection weights of 0.5 are used. The connection scheme is that illustrated in Figure 1A. Evaluation 1 is initially somewhat ambivalent, with a 0.5 activation on the positive pole (unit 1) and a 0.1 activation on the negative pole (unit 2). Evaluation 2 is initially slightly negative, with an activation of 0.1 on the negative pole (unit 4) and 0 on the positive pole (unit 3).

Activation updates over the first two time cycles of one particular run for this network are shown in Table 2. By default there are four unit updates per time cycle, the number of units in the network. Within each time cycle, the units to be updated are randomly selected. As it happens, unit 2 is selected for the first update. The current activation of this unit is 0.1 and the net input to the unit is −0.25, as specified in Equation 5, before being scaled by the resistance parameter. Again, the net input is computed as the sum of products of the activations on sending units and the connection weights. In the present case, there are only three such nonzero products: from unit 1 (activation of 0.5 × weight of −0.5), from unit 4 (activation of 0.1 × weight of 0.5), and from unit 2 itself (activation of 0.1 × weight of −0.5). Summing these three products yields an unscaled net input of −0.25. Multiplying by the resistance scalar of 0.5 yields −0.125. Because this net input is negative, update Equation 4 applies. Equation 4 requires a computation of the distance of the current activation of 0.1 from the floor activation of 0; this distance is 0.1. The net input is multiplied by this distance and then added to the current activation of the unit, yielding an updated activation of 0.088.

The next unit randomly selected for updating is unit 4. Readers may continue these computations, illustrated in Table 2, to obtain a feel for activation updates. Such updates typically continue until the network settles into a steady state as indexed by the fact that unit activations are no longer changing very much or, equivalently, that overall consonance or dissonance is no longer changing very much.

Simulations

With more than 1,000 published entries in the cognitive dissonance literature (Cooper & Fazio, 1984; Thibodeau & Aronson, 1992), there is considerable choice in deciding what experiments to simulate. Our strategy has been to simulate at least one experiment within each of the principal research paradigms and subparadigms of cognitive dissonance theory. It is generally acknowledged that the major, highly reliable experimental paradigms in this literature are those of insufficient justification and free-choice. Within the insufficient justification paradigm, there are three somewhat distinct lines of research, involving prohibition, initiation, and forced compliance. The free-choice paradigm has not tended to generate multiple, distinct lines of research.

Insufficient Justification Paradigms

The insufficient justification paradigm deals with situations in which subjects engage in some counterattitudinal action with rather little justification. Dissonance theory predicts that the less the justification for the behavior, the greater the dissonance and, at least when it is difficult to retract one's action, the more people will be motivated to change their attitudes so as to provide additional justification for their action.

As just noted, there have been three principal paradigms of insufficient justification research (Lepper, 1983): prohibition, initiation, and forced compliance. The classic example of a prohibition study is the forbidden toy experiment, in which children are forbidden to play with an attractive toy under either mild or severe threat of punishment from an adult experimenter (Aronson & Carlsmith, 1963). Somewhat paradoxically, the children devalued the forbidden toy more under mild than under severe threat. The classic initiation experiment demonstrated that people initiated into a boring group liked the group better after having undergone a severe than after having undergone a mild initiation (Aronson & Mills, 1959). Finally, the original forced compliance experiment involved inducing

Table 2
Activation Updates Over the First Two Time Cycles for the Illustrative Two-Cognition Network Specified in the Text

Unit number (randomly selected)	Current unit activation	Net input to unit	Scaled by unit resistance	Distance to floor or ceiling	Updated unit activation
2	0.100	−0.250	−0.125	0.100	0.088
4	0.100	−0.256	−0.128	0.100	0.087
2	0.088	−0.250	−0.125	0.088	0.077
1	0.500	−0.332	−0.166	0.500	0.417
3	0.000	0.127	0.063	1.000	0.063
4	0.087	−0.245	−0.123	0.087	0.076
2	0.077	−0.240	−0.120	0.077	0.067
2	0.067	−0.236	−0.118	0.067	0.059

subjects to voice a belief contrary to their own for either a large or a small amount of money (Festinger & Carlsmith, 1959). Again, somewhat paradoxically, smaller payments produced more attitude change.

Such counterintuitive predictions and findings contributed to the view that dissonance reduction was a somewhat exotic and irrational process compared to more conventional theories of conflict, decision, reinforcement, or rational choice. Dissonance results, especially within the insufficient justification paradigm, consistently seemed to go against traditional common sense. To proponents, such nonobvious findings were seen as evidence of the theory's ability to generate unique predictions; to opponents, such findings were seen as presumptive evidence of the theory's inapplicability to everyday life. Both agreed, however, that such results seemed to set dissonance apart from other psychological theories.

Within each subcategory of insufficient justification research, there remains a considerable choice of representative experiments to simulate. Our preference was to neglect both the original and the most modern studies in favor of second-generation experiments that effectively ruled out many of the alternative explanations that plagued the classic experiments yet remained focused on basic dissonance issues. Typically, these second-generation experiments involve a two-way interaction that is somewhat more challenging to simulate than the simple main-effects characteristic of the original experiments.

Prohibition. In the first of the insufficient justification studies to examine the consequences of prohibiting a desired action, nursery school children were forbidden to play with an attractive toy under either mild or severe threat of punishment (Aronson & Carlsmith, 1963). Both the mild and severe threats were carefully designed, however, to be sufficient to prevent the children from playing with the desirable toy during a play period in which the experimenter was absent from the room. In subsequent ratings, the children derogated the forbidden toy more under mild threat than severe threat. The original theoretical explanation was that the children had committed themselves to the dissonant behavior of not playing with the desirable toy. Because dissonance is greater when there are fewer cognitions to support the behavior, there was more dissonance in the mild threat condition than in the severe threat condition. Because the counterattitudinal behavior of not playing with the toy could not be retracted, dissonance was reduced by derogating the forbidden toy. The more the dissonance, the more the derogation. This basic result has been replicated in perhaps a dozen subsequent studies (e.g., Lepper, 1973; Pepitone, McCauley, & Hammond, 1967).

Close on the heels of the publication of Aronson and Carlsmith's classic study (1963), a number of alternative explanations of these findings were offered. These included the notion that severe threat focused more attention on the toy or made it seem more desirable, the idea that the experimenter appeared more likeable or more reasonable in the mild threat condition, and the possibility that the mild threat was seen as more credible.

To rule out these and other related alternative explanations, Freedman (1965) added surveillance conditions to the experiment. In the surveillance conditions, the experimenter stayed in the room while the child played. In these surveillance conditions, the same threats were used, but temptation, and thus dissonance, was lowered by the continued presence of the experimenter. Actual play with the previously forbidden toy 5 weeks later, in the absence of the experimenter or any continued prohibition, indicated greater derogation in the mild than in the severe threat conditions only when there was no surveillance. Results from Freedman (1965) are plotted in Figure 2A, in terms of the proportions of children playing with the forbidden toy 5 weeks later. Fewer children played with the forbidden toy in the mild than in the severe threat conditions only when there had been no surveillance. When children remained under surveillance, the effect of severity of threat was negligible. The results clearly support the dissonance explanation against the various alternative explanations.

Our simulation focused on this second-generation experiment by Freedman (1965).[6] Specifications of constraint satisfaction networks for the four conditions of this simulation are presented in Table 3. As in all of the insufficient justification networks, there are three relevant cognitions, concerning a behavior, a justification, and an evaluation. Following mapping relation 1, two units are used to encode each of the three cognitions: evaluation, threat, and play, making a six-unit network. Initially, the toy is given a high positive evaluation to reflect its desirability, *play with the toy* is given a high negative evaluation because it was not done, and the amount of threat is either low or high to represent mild or severe threats, respectively. All of this is in conformity with mapping relation 1. For the surveillance conditions of this simulation, the impact of both threats was scaled up by a multiplier in the spirit of the update rule specified in Equation 3:

$$\text{new\_threat} = \text{old\_threat} + (0.5 \times [1 - \text{old\_threat}]) \quad (8)$$

This made the value of threat 0.75 in the severe threat–surveillance condition and 0.55 in the mild threat–surveillance condition. These modifications, in accord with mapping relation 2, reflect the idea that surveillance enhances the value of both threats in accordance with the way that activations normally change.

Following mapping relation 2, connections across different cognitions reflect assumed causal implications among the cognitions. For the Freedman simulation, there were positive connections between toy evaluation and play (the better liked the toy, the more it would be played with), positive connections between toy evaluation and threat (the better liked the toy, the more threat would be required to prevent play), and negative connections between play and threat (the bigger the threat, the less the toy would be played with).

As a simulation begins, activations of units are updated in a random, asynchronous fashion. On each time cycle, n units are randomly selected and updated, using Equations 3, 4, and 5. By default, n is the number of units in the network, six in this simulation. This updating process implements mapping relation 4, concerning the motivation to reduce dissonance. Updating continued for 20 cycles because activation asymptotes were

[6] A preliminary simulation of Freedman (1965) was presented in Shultz and Lepper (1992).

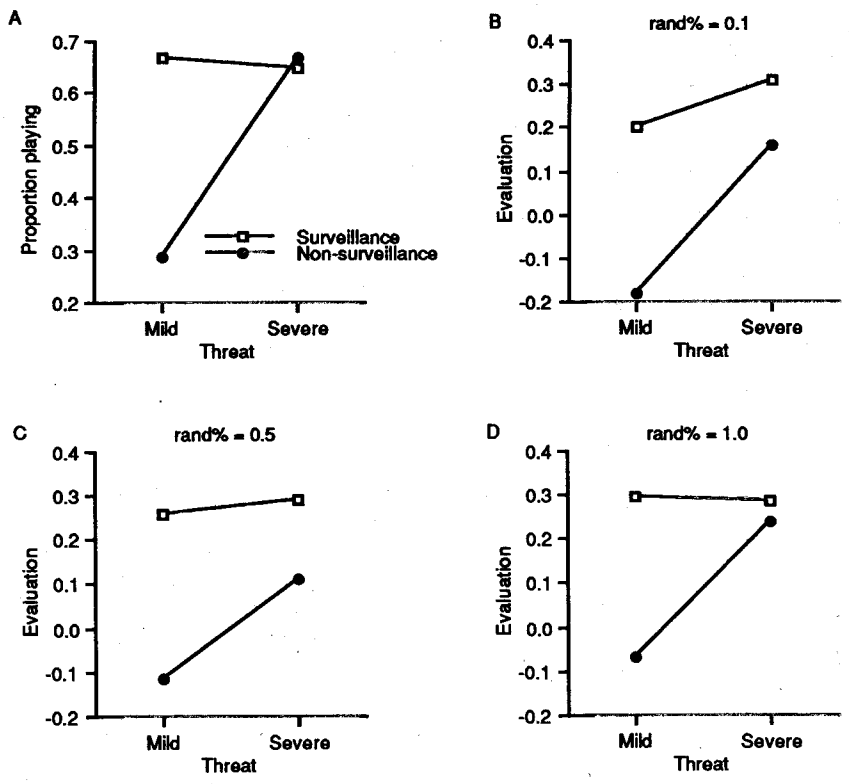

Figure 2. Human data from Freedman (1965), shown in Figure 2A, and simulation data at three levels of random parameter distortion (Figures 2B–2D). The human data are in terms of proportion of children playing with the previously forbidden toy in the nonforbidden session. Simulation data are in terms of net activation of the *evaluation of the toy* cognition after 20 update cycles.

reached within that period. We ran 20 networks in each condition at each of the three levels of rand%.

Mean evaluation of the forbidden toy after cycle 20 is shown in Figures 2B–2D for each of the three levels of rand%. Evaluation of the toy was computed as the difference between activation of the positive unit and the negative unit for the toy dimension. These evaluations were subjected to analyses of variance (ANOVAs) in which the presence of surveillance and severity of threat were the variables. Of primary interest was Freedman's (1965) predicted interaction between surveillance and severity of threat, $F(1, 76) = 97.73$, $p < .0001$, for rand% = 0.1; $F(1, 76) = 4.72$, $p < .05$, for rand% = 0.5; and $F(1, 76) = 3.54$, $p < .07$, for rand% = 1.0. In general, there was more derogation in the mild than in the severe condition, and this effect was much larger without surveillance than with surveillance. Statistically, this pattern weakens as randomization of parameters increases.

It is possible in our network models to examine the reduction of dissonance over cycles, as defined in Equation 6 and mapping relation 4. At present this is mainly of theoretical interest because there is no such direct measure of dissonance over time in human participants. Figure 3 contains plots of mean dissonance over cycles for four different simulations, all at rand% = 0.1. Figure 3A contains such data for the Freedman (1965) simulation. It shows the mild threat condition under nonsurveillance to start with the greatest amount of dissonance, but this dissonance is reduced dramatically over cycles. The other three conditions start with only moderate levels of dissonance that do not change much over cycles.

Differentiating among the four conditions of the Freedman experiment was accomplished only by manipulating the initial activations on the threat cognitions. Unlike the other insufficient justification experiments but like the free-choice experiments reported here, no connection weight changes were required. Also important to the Freedman results was the minus ceiling parameter of 0.5. A minus ceiling of 1 introduced a steep slope to the surveillance conditions line, with severe threat producing much higher evaluations than mild threat.

Initiation. Studies of insufficient justification through initiation grew out of the popular belief that, for example, the most popular fraternities seem to have the worst initiations, and the best of the armed services branches have the most demanding basic training. Such correlations were most often discussed as the result of a selection process wherein the less capable, attractive, or dedicated applicants were eliminated from possible membership.

A more counterintuitive explanation of these phenomena was that a severe initiation may itself increase liking for a group. This idea was first examined within the context of dissonance theory by Aronson and Mills (1959). Aronson and Mills induced dissonance in female university students by requiring them to pass an embarrassment test to participate in a discussion of sexual material in a group the students had previously

Table 3
Network Specifications for the Four Conditions of the Freedman Experiment

Condition	Cognition	Name	Type	+Activation	−Activation	Relation	Cause	Effect	Form
Nonsurveillance, mild threat	1	Evaluation	Evaluation	High	0	1	Evaluation	Play	Positive
	2	Play	Behavior	0	High	2	Evaluation	Threat	Positive
	3	Threat	Justification	Low	0	3	Threat	Play	Negative
Nonsurveillance, severe threat	1	Evaluation	Evaluation	High	0	1	Evaluation	Play	Positive
	2	Play	Behavior	0	High	2	Evaluation	Threat	Positive
	3	Threat	Justification	High	0	3	Threat	Play	Negative
Surveillance, mild threat	1	Evaluation	Evaluation	High	0	1	Evaluation	Play	Positive
	2	Play	Behavior	0	High	2	Evaluation	Threat	Positive
	3	Threat	Justification	0.55	0	3	Threat	Play	Negative
Surveillance, severe threat	1	Evaluation	Evaluation	High	0	1	Evaluation	Play	Positive
	2	Play	Behavior	0	High	2	Evaluation	Threat	Positive
	3	Threat	Justification	0.75	0	3	Threat	Play	Negative

volunteered to join. In the mild initiation condition, students read a list of mildly sex-related words to a male experimenter. In the severe initiation condition, they had to read a list of obscene words and a lurid passage from a novel to the same male experimenter. Then participants in each condition were to audit a supposed group discussion. The discussion had actually been tape recorded previously and was designed to be quite boring and banal. Following the discussion, the participants rated the group.

The dissonance theory prediction, and the empirical finding of the study, was that the worse the initiation, the better the group would be liked. Cognitions consonant with joining the group included the participant's favorable attitude to both the discussion topic and the group, for which she had volunteered. Cognitions dissonant with joining the group involved the embarrassment test, which was of course much worse in the severe condition than in the mild condition. Dissonance could be reduced by an increased evaluation of the group. Because dissonance was greater in the severe condition than in the mild condition, there should be more of an increase in liking of the group in the former condition.

At least five different alternative explanations were quickly offered for these findings. One suggested that the content of initiation and the discussion group were related. If the initiation had been arousing, this could have led to more interest in the group. In a related tack, perhaps these students did not know the meanings of at least some of the obscene words and wanted to join the group to discover what they were. Alternatively, they might have been intrigued by the obscene words and inferred that such words might eventually be discussed by the group.

There was also another explanation based on relief. The obscene material in the severely embarrassing test was followed by a very banal discussion. This sequence could have induced the arousal and then the reduction of anxiety, which then could have set the stage for a favorable evaluation of the group. Yet another explanation suggested that the participant might become dependent on the experimenter, more so in the severe than in the mild condition. This dependency might somehow mediate liking of the discussion group.

Still another idea was the afterglow hypothesis. All subjects had been told that they had passed the embarrassment test. In the severe condition, in which the test would probably be viewed as more difficult, participants might take more pride in their accomplishment, as compared with the mild condition. Finally, there was a contrast hypothesis that claimed that any experience following a negative experience would be experienced as pleasant. The worse the negative experience, the more positively the next experience is regarded.

Gerard and Mathewson (1966) neatly disposed of all of these alternative explanations by separating the content of the initiation experience from that of the group discussion and by adding noninitiation conditions with the same severe and mild levels of discomfort. They administered mild or severe electric shock, either as part of an initiation or as part of a second, unrelated "psychological experiment." Following this, participants heard a boring group discussion of cheating in university courses.

It was predicted from dissonance theory, and found, that participants who received severe shock liked the group better than did subjects who received mild shock, but only in the initiation

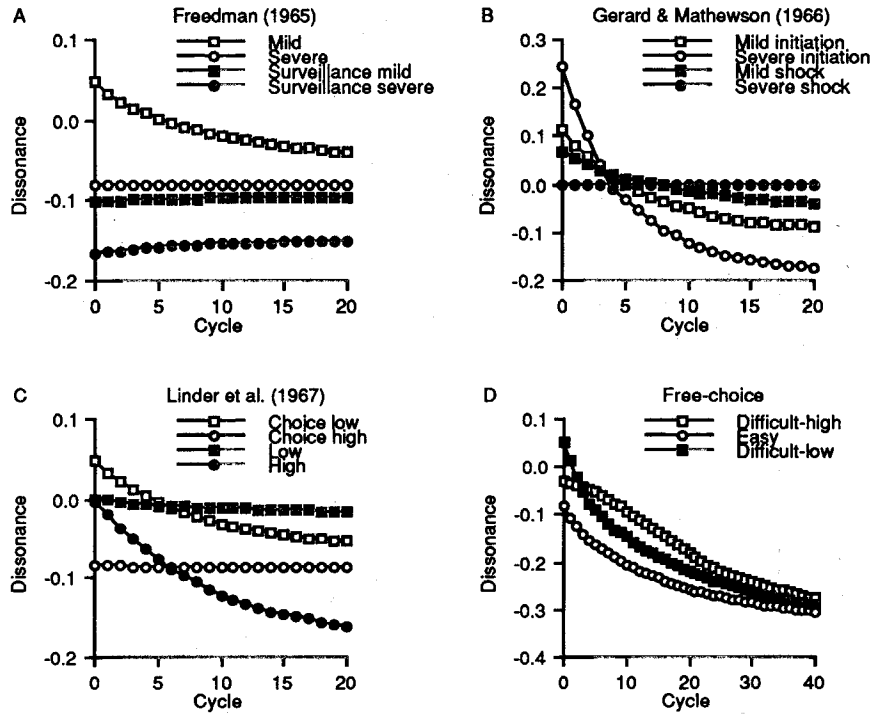

Figure 3. Mean dissonance over cycles at rand% = 0.1 for four simulations. Cycle 0 refers to initial dissonance, before any activation updates.

condition. It was found, but not predicted by dissonance theory, that the reverse trend held in the noninitiation condition, that is, the worse the shock, the less the group was liked. Without initiation, dissonance theory cannot predict a difference between mild and severe shock. The obtained interaction is presented in Figure 4A. A similar interaction was obtained for participants' ratings of the group discussion.

Network specifications for the four conditions of the Gerard and Matthewson experiment are presented in Table 4. Again there are three relevant cognitions and three relations among them. For the initiation condition, there is an excitatory relation between evaluation of the group and joining the group (the better you like the group, the more likely you are to join it). There is also an excitatory relation between evaluation of the group and degree of shock (you get what you pay for). The relation between degree of shock and joining the group is negative (a rising price lowers demand). Initial activations stem from application of mapping relation 1. Joining the group is given an initial positive value to reflect the fact that the participants did volunteer to join. Evaluation of the group is given an initial negative value to reflect the subsequent boring nature of the actual discussion. Shock levels are given positive initial values, reflecting the fact that all participants had just received shocks. High values were used for severe shock, and low values for mild shock.

Networks for the noninitiation condition were similar except that relations between shock level and joining the group were cut to 0; these dimensions are no longer causally related because the shock is not part of an initiation. Moreover, the relations between shock and evaluation of the group are changed from excitatory to inhibitory because being shocked no longer pays for the right to join the group; instead, the negative experience of being shocked adversely affects how one feels about the whole experimental session. These connection weight changes are consistent with the idea of following the causal relations among cognitions specified in mapping relation 2. As noted earlier, parameter values for initial activations and resistances are the same as those used in the previous, as well as succeeding, simulations.

Simulation results, in terms of evaluation of the group after cycle 20, are presented in Figures 4B–4D for three levels of rand%. These evaluations were subjected to ANOVAs in which the presence of initiation and severity of shock were the variables. Of primary interest was the predicted interaction between initiation and shock, $F(1, 76) = 945$, $p < .0001$, for rand% of 0.1; $F(1, 76) = 54.28$, $p < .0001$, for rand% of 0.5; and $F(1, 76) = 3.41$, $p < .07$, for rand% of 1.0. The dissonance effect holds only in the initiation condition; the more the shock, the better the group is liked. In the noninitiation condition, there is what could be called an annoyance effect; the more the shock, the less the rest of the session is appreciated. The constraint satisfaction networks thus make a more complete fit to the human data than does dissonance theory, which does not predict the annoyance effect.

Reduction of dissonance over cycles for this simulation is shown in Figure 3B, as defined in Equation 6 and mapping relation 4. It reveals that the relatively high level of dissonance in the severe initiation condition is substantially reduced over time cycles. There is some lesser amount of dissonance reduction in the mild initiation and mild shock conditions, but virtually none in the severe shock condition.

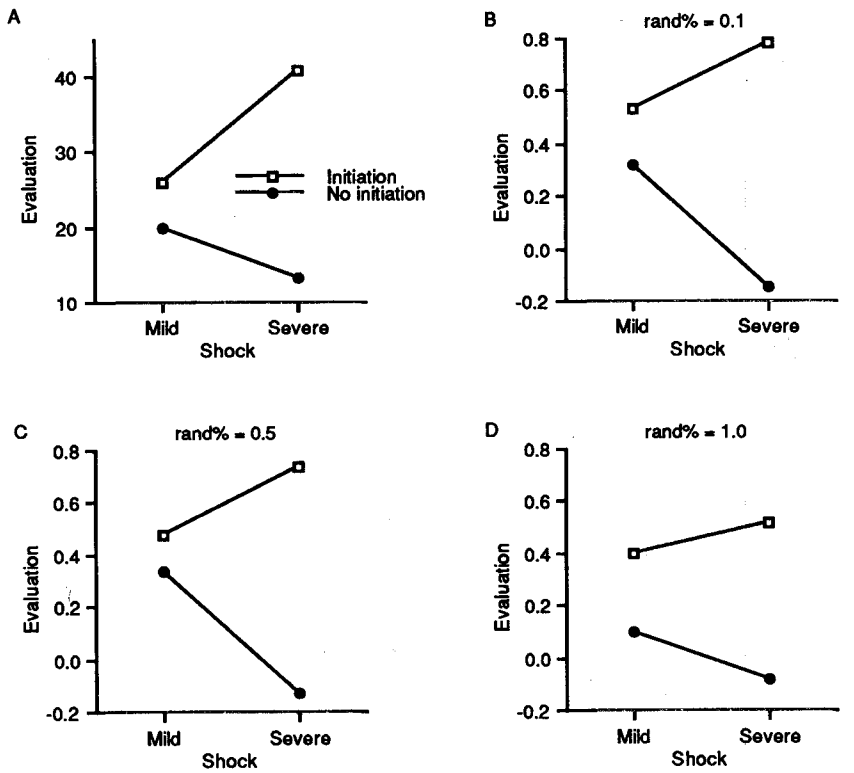

Figure 4. Human data from Gerard and Mathewson (1966), shown in Figure 4A, and simulation data at three levels of random parameter distortion (Figures 4B–4D). The human data are in terms of evaluation of the group participants. Simulation data are in terms of net activation of the *evaluation of the group* cognition after 20 update cycles.

Differentiating among the four conditions of the Gerard and Mathewson experiment required both initial activation and connection weight changes. Differentiating mild and severe shock was merely a matter of using low versus high initial activations on the shock cognition, respectively. Differentiating initiation versus noninitiation conditions required two changes in intercognition relations: the sign of the relation between the shock and evaluation cognitions, and the presence or absence of an inhibitory relation between the shock and join cognitions. To verify that both of these relation changes were necessary, we tried the noninitiation conditions in two additional ways: (a) shock–evaluation and shock–join relations both inhibitory, and (b) shock–evaluation relation facilitory and shock–join relation 0. The former produced adequate data fits, but the mild shock conditions were a bit too close together. Here, the noninitiation results were due to the inhibitory relation for shock–evaluation because that is all that differed from the initiation conditions. The latter (b) yielded severe shock greater than mild shock evaluations for noninitiation as well as for initiation, which is clearly wrong. Here the 0 relation for shock–join was responsible because it was the only difference. In summary, both relation changes are required to produce the correct results in the noninitiation conditions, namely, shock–evaluation relations inhibitory and shock–join relations 0.

Unlike the other simulations presented here, the Gerard and Mathewson simulations were decidedly more accurate with a negative cap value. With a cap of 0, the mild shock evaluations and the initiation evaluations were too high, giving zero slope to the initiation evaluation line. The proper interaction emerged with a cap of −0.5.

Forced compliance. The third major insufficient justification paradigm involves paying people to do something that is discrepant with their own attitudes. In the classic study of this sort, Festinger and Carlsmith (1959) gave university students an extremely dull task to do and then paid them either $1 or $20 to tell the next participant, who was actually a confederate of the experimenter, that the task was interesting and enjoyable. Subsequently, participants were asked to give their own personal opinion of the task. Those who had been paid $1 to lie rated the task higher than those who had been paid $20 to lie, as predicted by dissonance theory. Knowledge that the task is dull is dissonant with having said it was interesting and enjoyable. The payment is consonant with the lie, more so with the larger $20 payment. Subjects could reduce dissonance by increasing their personal evaluation of the task. They should do this more in the $1 condition than in the $20 condition because there is more overall dissonance to be reduced with the smaller payment than with the larger payment. Like many other findings in the insufficient justification paradigm, this result was considered counterintuitive. It seemed to contradict the notion from reinforcement theory that the strength of a reinforced behavior increases with the degree of reinforcement. In this case, the payment might

Table 4
Network Specifications for the Four Conditions of the Gerard and Mathewson Experiment

Condition	Cognition	Name	Type	+Activation	−Activation	Relation	Cause	Effect	Form
Mild initiation	1	Evaluation	Evaluation	0	High	1	Evaluation	Join	Positive
	2	Join	Behavior	High	0	2	Shock	Evaluation	Positive
	3	Shock	Justification	Low	0	3	Shock	Join	Negative
Severe initiation	1	Evaluation	Evaluation	0	High	1	Evaluation	Join	Positive
	2	Join	Behavior	High	0	2	Shock	Evaluation	Positive
	3	Shock	Justification	High	0	3	Shock	Join	Negative
Noninitiation, mild shock	1	Evaluation	Evaluation	0	High	1	Evaluation	Join	Positive
	2	Join	Behavior	High	0	2	Shock	Evaluation	Negative
	3	Shock	Justification	Low	0	3	Shock	Join	0
Noninitiation, severe shock	1	Evaluation	Evaluation	0	High	1	Evaluation	Join	Positive
	2	Join	Behavior	High	0	2	Shock	Evaluation	Negative
	3	Shock	Justification	High	0	3	Shock	Join	0

have constituted reinforcement for the behavior of giving a positive description of the task.

As was true of other insufficient justification results, alternative explanations soon emerged. One was based on evaluation apprehension, the idea that participants in psychology experiments may think they are being evaluated. In particular, in the Festinger and Carlsmith (1959) experiment, subjects may have felt that their honesty was being tested. To resist the temptation to lie for money might have resulted in a more favorable evaluation by the experimenter. Perhaps evaluation apprehension would have been higher with the $20 payment than with the $1 payment.

Other alternative explanations were based on deception. Subjects may have suspected that they were being deceived and, as a result, may have angrily resisted confirming what they perceived to be a test of the reinforcement hypothesis. Alternatively, subjects in the $20 condition may have found the payment far too large for the situation and may have inferred that the payment was actually designed, instead, to alter their beliefs.

These and other possible alternative explanations were effectively ruled out in a follow-up experiment by Linder, Cooper, and Jones (1967). Reasoning that dissonance effects would obtain only under free-choice, Linder et al. added no-choice conditions in which alternative factors, but not dissonance, would be expected to operate. Their college student subjects were asked, under either choice or no-choice conditions, to write a forceful essay supporting a ban on communist speakers on campus—a position with which these students strongly disagreed. They were paid either $0.50 or $2.50 to write this counterattitudinal essay. The predicted crossover interaction occurred such that, in the choice condition, banning communist speakers was favored more with low pay than with higher pay, whereas in the no-choice condition, banning communist speakers was favored more with higher pay than with low pay. These results are presented in Figure 4A.[7] Notice that dissonance theory per se predicts only the results obtained in the choice condition. Dissonance theory is not really applicable to the no-choice condition, in which other processes are assumed to operate.

Network specifications for the four conditions of the Linder et al. experiment are presented in Table 5. As with other insufficient justification simulations, there are three cognitions (concerning a behavior, a justification, and an evaluation) with three relations among them. These relations reflect mapping relation 2. In the choice condition, there were excitatory relations

[7] Extensive subsequent research on forced compliance (e.g., Calder, Ross, & Insko, 1973; Collins & Hoyt, 1972; Cooper & Fazio, 1984; Wicklund & Brehm, 1976) has replicated and extended the findings of Linder et al. (1967). As a whole, the literature suggests three related preconditions necessary for forced compliance procedures regularly to produce dissonance effects. These preconditions, typically summarized as personal-responsibility-for-(negative)-consequences, include a perception that one had a choice not to engage in the counterattitudinal action, a feeling of personal responsibility for one's action, and a belief that one's action will have negative consequences. It is worth noting that the same principles used to simulate Linder et al.'s (1967) findings concerning choice could also be used to model the effects of responsibility and consequences.

Table 5
Network Specifications for the Four Conditions of the Linder et al. Experiment

Condition	Cognition	Name	Type	+Activation	−Activation	Relation	Cause	Effect	Form
Choice, low payment	1	Attitude	Evaluation	0	High	1	Attitude	Essay	Positive
	2	Essay	Behavior	High	0	2	Pay	Essay	Positive
	3	Payment	Justification	Low	0	3	Attitude	Payment	Negative
Choice, high payment	1	Attitude	Evaluation	0	High	1	Attitude	Essay	Positive
	2	Essay	Behavior	High	0	2	Pay	Essay	Positive
	3	Payment	Justification	High	0	3	Attitude	Payment	Negative
No choice, low payment	1	Attitude	Evaluation	0	High	1	Attitude	Essay	0
	2	Essay	Behavior	High	0	2	Pay	Essay	Positive
	3	Payment	Justification	Low	0	3	Attitude	Payment	Positive
No choice, high payment	1	Attitude	Evaluation	0	High	1	Attitude	Essay	0
	2	Essay	Behavior	High	0	2	Pay	Essay	Positive
	3	Payment	Justification	High	0	3	Attitude	Payment	Positive

between the attitude toward banning communist speakers and the writing of the essay (the more one supports this position, the more likely one would express this position in writing) and between the essay and payment (you get what you pay for). The relation between the attitude toward the ban and payment was negative, reflecting the idea that the more favorable one's attitude the less one would need to be paid to write an essay at some particular level of support.

Initial unit activations correspond to mapping relation 1. The initial attitude was specified as high negative to reflect the relatively liberal attitudes of these university students. Writing the essay was given a high positive initial activation because the essay was in fact written. Payment for writing the essay was either high or low depending on the payment condition.

The networks for the no-choice condition were identical to those for the choice condition, with two exceptions consistent with mapping relation 2. First, the relation between attitude and essay was changed to 0. Without a choice, there should be no causal relation between one's attitude and the writing of an essay on the same topic. Second, the relation between attitude toward banning and payment for the essay was changed to positive. The more one is paid, the more positively one should evaluate the whole session, including the topic of the essay. We favor describing this as a *mood,* rather than a reinforcement, *effect* but relevant psychological evidence is not yet definitive. It is analogous to the argument that being shocked for no good reason leads participants to lower their evaluation of being in the Gerard and Mathewson experiment.

Simulation results, in terms of mean attitude toward banning after cycle 20, are presented in Figures 5B–5D at three levels of rand%. These evaluations were subjected to ANOVAs in which the presence of choice and amount of payment were the variables. The interaction between choice and payment was statistically reliable at each level of rand%: $F(1, 76) = 1791, p < .0001$, for rand% of 0.1; $F(1, 76) = 57.71, p < .0001$, for rand% of 0.5; and $F(1, 76) = 8.02, p < .001$, for rand% of 1.0, although the crossover almost disappeared at rand% 1.0. These interactions reflect the presence of a dissonance effect under choice and a mood effect under no-choice, even though there were differences in the precise rank orderings of the conditions. This supports the view that our consonance constraint satisfaction model is more general than dissonance theory, which cannot predict the mood (or alternatively, reinforcement) effect.

Mean dissonance over cycles, as defined in Equation 6 and mapping relation 4, at rand% = 0.1, is presented in Figure 3C. This plot reveals relatively high dissonance in the choice, low-pay condition, which is greatly reduced over time. Dissonance also drops in the no-choice, high-pay condition, although from a lesser height.

Differentiating among the four conditions of the Linder et al. experiment required both initial activation and connection weight differences. The difference between high and low payments was conveyed simply by using high and low initial activations, respectively. Differentiating between the choice and no-choice conditions was a matter of two connection weight changes: the presence or absence of a facilitory relation between the attitude and essay cognitions, and the sign of the attitude–payment relation. To verify that both of these connection weight changes were necessary, we tried doing the no-choice conditions

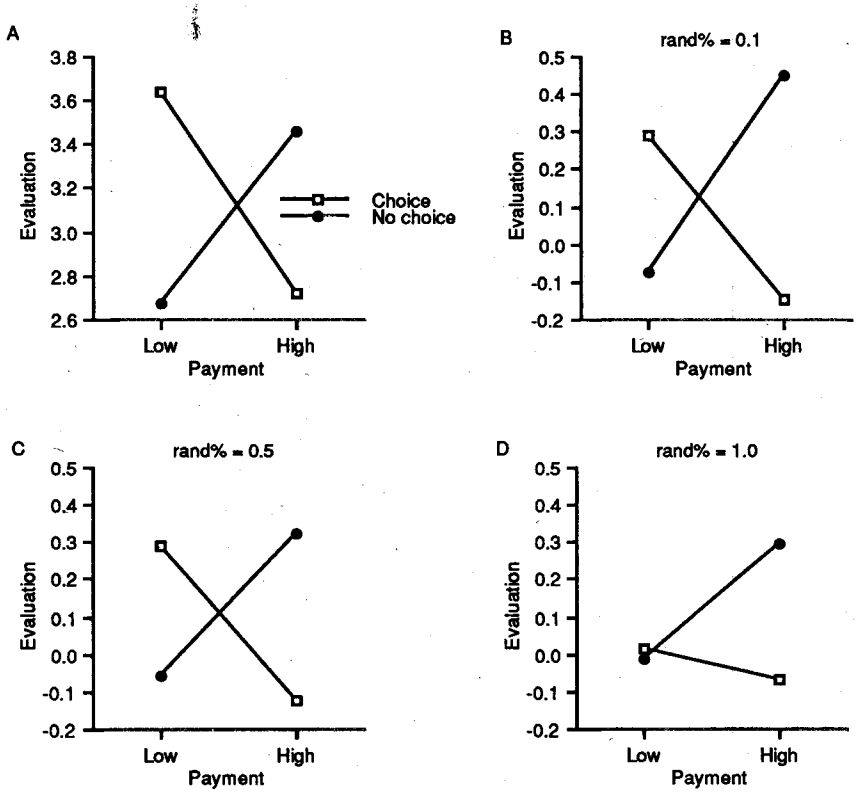

Figure 5. Human data from Linder et al. (1967, Experiment 2), shown in Figure 5A, and simulation data at three levels of random parameter distortion (Figures 5B–5D). The human data are in terms of evaluation of the counterattitudinal position. Simulation data are in terms of net activation of the evaluation of the counterattitudinal cognition after 20 update cycles.

in two additional ways: (a) attitude–essay and attitude–payment relations both positive, and (b) attitude–essay relation 0 and attitude–payment relation negative. Method (a) gave correct slopes but no crossover, that is, the evaluations under no-choice were both greater than under choice. Here the positive attitude–pay relation was responsible because that was the only difference from the choice conditions. Method (b) yielded higher evaluations for low payment than for high payment under no-choice as well as under choice. The superiority of low payment was somewhat muted for no-choice, producing an interaction, but not enough for the required crossover. The attitude–essay relation of 0 was responsible for these results because that was the only difference from choice conditions. In summary, both connection weight changes are required to produce the correct psychological results in the no-choice conditions, namely, attitude–essay relations of 0 and positive attitude–pay relations.

As with all experiments except the Gerard and Mathewson experiment, data fits were better for the Linder et al. experiment with a minus ceiling of 0.5 than with a minus ceiling of 1. In simulating Linder et al., a minus ceiling of 1 made the low-payment evaluations too close together across the choice and no-choice conditions.

Thus, the three target experiments on insufficient justification were effectively simulated by the consonance constraint satisfaction model. The fact that these three experiments are representative of many other studies in their respective sectors suggests that the consonance model could capture those other results too. With insufficient justification behind us, we can move on to the other highly reliable major dissonance paradigm, that of free-choice.

Free-Choice Paradigm

Choosing between alternatives creates cognitive dissonance because of the fact that the chosen alternative is never perfect and the rejected alternative often has desirable aspects that are necessarily foregone when an irreversible choice is made. Once a choice is made, dissonance can be reduced by viewing the chosen object as more desirable and by viewing the rejected object as less desirable. Such dissonance reduction will further separate the alternative choices in terms of their desirability. Magnitude of dissonance would be greater the closer the alternatives are in desirability before the choice is made. The closer the alternatives are in their initial desirability, the more difficult an exclusive choice between them is.[8] The greater the dissonance created by the choice, the more the increase in separation between the alternatives after the choice has been made.

[8] Note, of course, that there would be no dissonance if the two alternatives were identical except in magnitude (e.g., a choice between $1 and $2). Dissonance depends on the presence of qualitative differences between alternatives.

The classic free-choice experiment required female university students to rate eight different small appliances (Brehm, 1956). They were then given either a difficult choice (i.e., between two objects that they had rated high) or an easy choice (i.e., between one object they had rated high and another object they had rated low) of one item to take home in payment for their services. Then they rated the objects again. Amount of separation was measured by subtracting the first rating from the second rating for each of the two objects. Even though dissonance theory predicted greater separation in the difficult choice condition than in the easy choice condition, most of the actual separation obtained was due to a relatively large decrease in the value of the rejected alternative in the difficult choice condition. This pattern is graphed in Figure 6A in terms of mean evaluation change.

We refer to the difficult choice condition in Brehm's experiment as difficult–high because both alternatives had high initial evaluations. In our simulations, we added a difficult–low condition, which to our knowledge had not been used in previous free-choice experiments.[9]

Network specifications for the three conditions of the free-choice experiments are presented in Table 6. As in all of our dissonance simulations, there are three cognitions and three relations among them. However, in the case of free-choice, two of the cognitions are of the evaluation type (one for each choice alternative) and one refers to a behavior (the decision). There was a positive relation between the decision and evaluation of the chosen alternative to reflect the fact that it was chosen, and there was a negative relation between evaluation of the two alternatives to reflect the fact that they are in competition. Initial unit activations were consistent with mapping relation 1. Initial activation of the decision was high, reflecting a definite public decision. Initial activations for the chosen and rejected alternatives varied with particular choice conditions, as specified in mapping relation 1. These evaluations varied symmetrically around 0 to allow for full use of the activation ranges. This proved to be important in the context of computing evaluation change scores. For the difficult–high condition, evaluation of the chosen alternative was given an initial activation of 0.3 and evaluation of the rejected alternative an initial activation of 0.2. For the easy condition, initial evaluation activations were 0.3 for the chosen alternative and −0.3 for the rejected alternative. For the difficult–low condition, the initial evaluation activations were −0.2 for the chosen alternative and −0.3 for the rejected alternative.

Free-choice networks were run for 40 cycles because it was clear that they did not reduce dissonance to asymptotic values until about then.

Mean difference scores (reevaluation minus initial evaluation) after 40 cycles are plotted in Figures 6C–6E for three different levels of rand%. Evaluation was computed as the difference in activation between the positive and negative units. These evaluations were subjected to ANOVAs in which the nature of the choice served as a between-networks variable and choice alternative served as a within-network variable. The interaction between alternative and condition was significant for rand% of 0.1, $F(2, 57) = 301.49$, $p < .0001$; and 0.5, $F(2, 57) = 21.44$, $p < .0001$, but not for rand% of 1.0, $F(2, 57) = 1.94$. Figures 6C–6E show that most of the action was produced by a decrease in evaluation of the rejected alternative in the difficult–high condition and an increase in evaluation of the chosen alternative in the difficult–low condition. As parameter randomization increases, the interaction weakens statistically but the pattern of evaluation change remains fairly constant.

It is apparent that this interaction matches Brehm's (1956) human data (Figure 6A) rather precisely. Considering only the difficult–high and easy conditions used in Brehm's experiment, most of the action in both the simulation and the human data was due to the drop in evaluation of the rejected alternative. These simulation results thus fit Brehm's (1956) human data more precisely than does dissonance theory, which merely predicts a larger separation of the alternatives following a difficult choice than following an easy choice.

Use of the new difficult–low condition in the simulation provides some predictions of the consonance model. In this difficult–low condition, most of the action is created by the rise in evaluation of the chosen object. Moreover, the rise in evaluation of the chosen object in the difficult–low condition appears greater than the fall in evaluation of the rejected object in the difficult–high condition. This was assessed by an unpaired t test, comparing the negative of the difficult–high rejected scores to the difficult–low chosen scores. This two-tailed test was significant at each level of rand%: $t(38) = 28.47$, $p < .0001$, for rand% of 0.1; $t(38) = 6.70$, $p < .0001$, for rand% of 0.5; and $t(38) = 2.81$, $p < .01$, for rand% of 0.1.

These predictions were tested in a free-choice experiment with 13-year-olds (Shultz, Léveillé, & Lepper, 1995). Participants rated eight attractive posters, made a choice between two of them, and then rated the posters again. Choice conditions included difficult–high, easy, and difficult–low. The results, in terms of mean evaluation change, are presented in Figure 6B. A contrast regression F test, reflecting the interaction pattern predicted by the simulations (Figures 6C–6E), had weights of −1 for the easy and difficult–low rejected cells, −3 for the difficult–high rejected cell, +1 for the difficult–high and easy chosen cells, and +3 for the difficult–low chosen cell, $F(1, 153) = 79.06$, $p < .001$.[10] To ensure that the predicted interactions held for both the difficult–high and difficult–low conditions, we examined separate contrasts for each in comparison to the easy condition. In one contrast, weights were +2 for the difficult–high and easy chosen cells, −3 for the difficult–high rejected cell, −1 for the easy rejected cell, and 0 for the difficult–low cells, $F(1, 153) = 20.50$, $p < .001$. In the other contrast, weights were +1 for the easy chosen cell, +3 for the difficult–low chosen cell, −2 for the easy and difficult–low rejected cells, and 0 for the difficult–high cells, $F(1, 153) = 54.34$, $p < .001$.

Yet another way to examine these data, consistent with the original Brehm (1956) study, is by determining which of the six means differs significantly from the theoretical mean of 0, representing no change. Dunnett's (1955) technique for comparing a number of treatment means against a control mean was modified to use a theoretical control mean. Application of this technique revealed that only the means for the difficult–high rejected and difficult–low chosen conditions differed from

[9] A preliminary simulation of Brehm (1956), using the difficult–high and easy choices only, was presented in Shultz and Lepper (1992).

[10] The sum of such contrast weights must be zero.

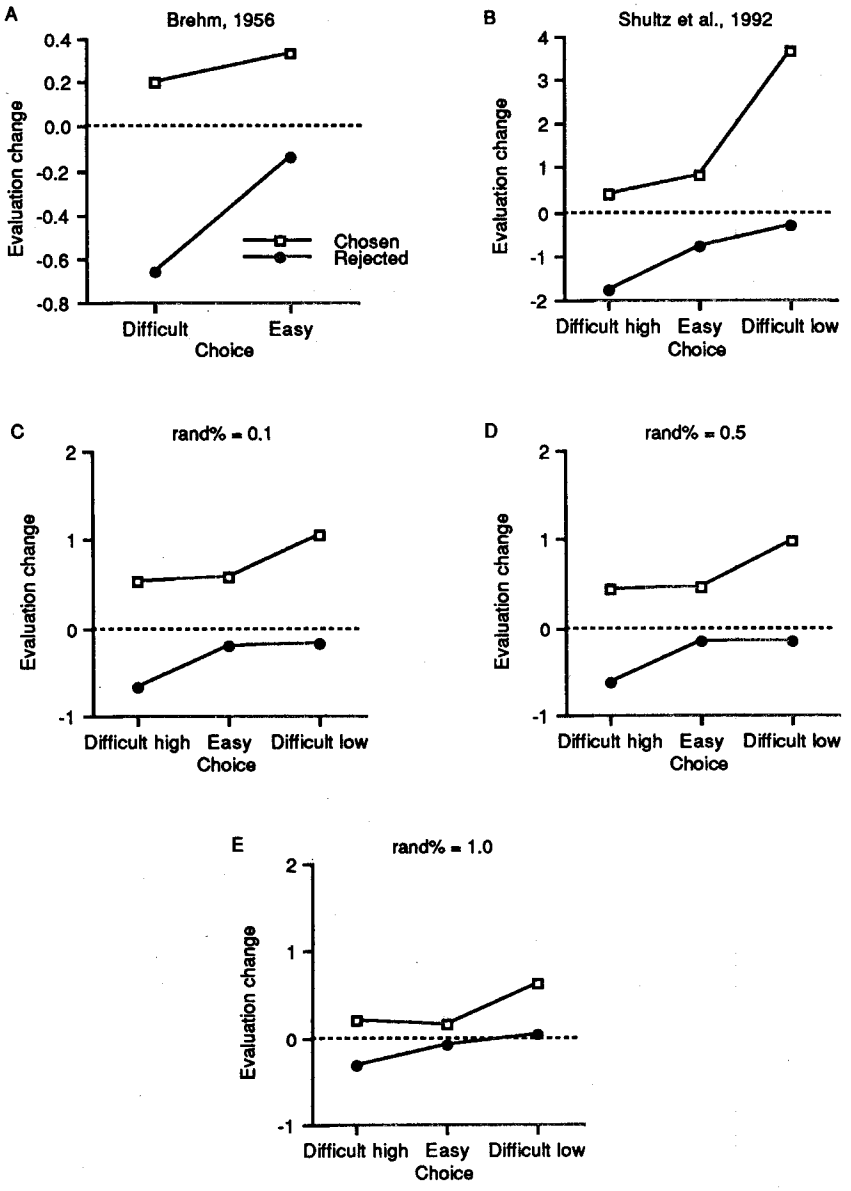

Figure 6. Human data on evaluation change in free-choice (Figures 6A–6B) and simulation data at three levels of random parameter distortion (Figures 6C–6E). Simulation data are after 40 update cycles.

0, $p < .01$. None of the other four means in Figure 6B differed from 0, $p > .05$.

As with the simulations, negatives of the difficult–high rejected scores were less than the difficult–low chosen scores, $t(50) = 2.97, p < .01$. In other words, the decrease in evaluation of the rejected object in the difficult–high condition was less than the increase in evaluation of the chosen object in the difficult–low condition.

Thus, the predictions from our consonance constraint satisfaction model of free-choice were confirmed. Again, the subtleties of this interaction are not predicted by dissonance theory, which only predicts more separation of the evaluations of the chosen and the rejected objects in the difficult conditions than in the easy condition.

The interaction between choice condition and alternative, in both the simulations and the human data, can be understood in terms of changing the evaluation of the alternative that has the most room to move in a given direction. For the chosen alternative, the direction of evaluative movement is up, and there is more room to move up in the difficult–low condition, in which the chosen object is not highly evaluated to begin with, than in the other two conditions, where it starts out with relatively high evaluation. For the rejected alternative, the direction of movement is down, and there is more room to move in that direction in the difficult–high condition, in which the rejected alternative starts with a high evaluation, than in the other two conditions, in which it starts low. Note that this is not to attribute the pattern of evaluation change to a mere statistical phenomenon

Table 6
Network Specifications for the Three Conditions of the Free-Choice Experiments

Condition	Cognition	Name	Type	+Activation	−Activation	Relation	Cause	Effect	Form
Difficult choice, high initial evaluation	1	Chosen	Evaluation	0.3	0	1	Decision	Chosen	Positive
	2	Rejected	Evaluation	0.2	0	2	Decision	Rejected	0
	3	Decision	Behavior	High	0	3	Chosen	Rejected	Negative
Easy choice	1	Chosen	Evaluation	0.3	0	1	Decision	Chosen	Positive
	2	Rejected	Evaluation	0	0.3	2	Decision	Rejected	0
	3	Decision	Behavior	High	0	3	Chosen	Rejected	Negative
Difficult choice, low initial evaluation	1	Chosen	Evaluation	0	0.2	1	Decision	Chosen	Positive
	2	Rejected	Evaluation	0	0.3	2	Decision	Rejected	0
	3	Decision	Behavior	High	0	3	Chosen	Rejected	Negative

known as regression to the mean. Regression to the mean is the tendency for both high and low scores to move toward the mean when reassessed under conditions of substantial measurement error.

The Shultz et al. (1995) experiment also included control conditions in which participants merely rated the posters twice, without making an intervening choice. A regression to the mean explanation would predict that mean change scores in these control conditions would resemble those found in the experimental choice conditions, that is, a large increase in evaluation of the better alternative in the difficult–low condition and a large decrease in evaluation of the lesser alternative in the difficult–high condition. Mean change scores for the control conditions were nothing like those in the choice conditions, hovering around 0 and thus demonstrating that the obtained interaction shown in Figure 6B could not be due merely to regression to the mean effects.

Mean dissonance reduction over cycles, reflecting Equation 6 and mapping relation 4, at rand% = 0.1, is presented in Figure 3D. Initial dissonance is greater in the difficult–low condition than in the difficult–high condition, which in turn has greater initial dissonance than the easy condition. Over cycles, dissonance decreases to about equal levels in all three conditions. Dissonance theory would have predicted more initial dissonance for a difficult–high choice than for an easy choice. Simulation with the consonance model suggests that there is even more initial dissonance in the difficult–low condition. Apparently, being forced to choose among disliked alternatives is especially dissonance arousing. Classical dissonance theory would appear to have no particular prediction to make about relative amounts of dissonance in difficult–low versus difficult–high choices. Unless perhaps dissonance theory could be stretched to predict more dissonance in the difficult–high than in the difficult–low condition because the initial low evaluations of the alternatives in the former condition may decrease the importance of the choice. This finding of greater dissonance in the difficult–low condition than the difficult–high condition can be used to explain why evaluation of the chosen object in the difficult–low condition increases more than evaluation of the rejected object in the difficult–high condition decreases. In brief, more dissonance leads to more evaluation change.

As in the Freedman experiment, differentiating among conditions in free-choice experiments required only manipulating initial activations, in this case on the evaluation cognitions. To capture the subtleties of the human data, it was important that these initial activations be symmetrical around 0, that is, the best liked objects had positive net evaluations and the worst liked objects had negative net evaluations. This symmetry around 0, coupled with the −ceiling parameter of 0.5, allowed evaluations of the chosen objects to increase more than evaluation of the rejected objects decreased. Other simulations indicated that evaluation of the rejected objects decreased too far below 0 with a −ceiling of 1.

Discussion

Simulation results with the consonance model matched psychological findings from the insufficient justification and free-choice paradigms of cognitive dissonance theory. In the case

of insufficient justification through prohibition, the consonance network produced the same interaction between surveillance and severity of threat found by Freedman (1965). There was more derogation of the forbidden toy with mild than with severe threat only under nonsurveillance.

For insufficient justification from initiation, the consonance network yielded the interaction between initiation and severity reported by Gerard and Mathewson (1966). When shock was part of the initiation procedure, severe shock produced more liking for the group than did mild shock. But when shock was unrelated to the initiation procedure, severe shock produced less liking for the group than did mild shock. Dissonance theory can predict only the effect obtained under initiation, not the annoyance-based reversal found with noninitiation.

For insufficient justification from attitude-discrepant behavior, the consonance network mimicked the crossover interaction between choice and payment found by Linder et al. (1967). Under choice, attitude change was greater with low than with high payment. But with no-choice, the results were reversed, with more attitude change under high payment than under low payment. Dissonance theory is able to predict the results in the choice condition but does not predict the mood or incentive effect in the no-choice condition.

In the free-choice paradigm, consonance network simulations captured the findings reported by Brehm (1956). As in the human experiment, the locus of most of the action was in the reevaluation of the rejected alternative in the difficult choice condition. In contrast, dissonance theory only predicted that the postchoice separation between the alternatives would be greater for difficult choices than for easy choices. Simulation predictions for close but lowly evaluated alternatives were supported in a new experiment (Shultz et al., 1995). Here, most of the change was the result of an increase in the evaluation of the chosen object in the difficult–low condition. For both simulations and human data, this increase in evaluation of the chosen object in the difficult–low condition was greater than the decrease in evaluation of the rejected object in the difficult–high condition.

In many of these cases, the consonance network simulations provide better coverage of the psychological data than does cognitive dissonance theory. These superior fits of the consonance network derive from the capacity of constraint satisfaction models to deal with variables other than those unique to cognitive dissonance theory and from the increased precision that is inherent to these computational formulations. Among the nondissonance effects captured by the consonance network were the annoyance effect in initiation studies, the mood effect in attitude-discrepancy studies, and the locus of change effects in free-choice experiments. Consonance constraint satisfaction is clearly more general than classical dissonance theory.

Network Design

Connection weights, caps, resistances, and initial activations assumed standard, randomized values across experiments, and each experiment was simulated within a generic consonance network that included three cognitions of three types: behaviors, justifications, and evaluations. Particular instantiations of this network were designed by using a standard set of constraints obtained by a formal mapping of dissonance theory to constraint satisfaction techniques: (a) representation of each key cognition in the experiment by a pair of negatively connected units, (b) use of connection weights to represent causal implications among the cognitions, (c) computation of total dissonance (or consonance) across the entire network of cognitions, (d) use of activation update rules that encouraged gradual reduction of dissonance (equivalently, increase in consonance), and (e) use of a resistance parameter to bias networks in favor of particular modes of dissonance reduction. Evaluations had less resistance to change than did beliefs about behaviors and justifications.

The fact that consonance networks need to be designed in a certain, principled way to capture dissonance phenomena shows that the issues being addressed by the simulations are not so self-evident or overdetermined that any constraint satisfaction model would suffice. Instead it suggests that the consonance model may have focused on the critical variables.

Initial activations and dissonance. All of these simulations began with some units having initial, nonzero activation values. It is more conventional for constraint satisfaction programs to start all units at zero activation and then to provide some of the units with external inputs. Activations then gradually build up from zero as a function of both external input and activation that is internal to the network. This is true even of many models of social phenomena such as balance, dissonance, and attitude change (Read & Miller, 1994; Spellman et al., 1993). Our pilot results showed that this conventional technique was not appropriate for our model of cognitive dissonance phenomena because it yielded results indicating a gradual increase in consonance but no dissonance. With initial activations of zero, there is no dissonance because triple products including any zeros will be zero. Because of the update formulas (Equations 3–5), consonance gradually increases until asymptote. Dissonance will never occur as the network moves from zero to nonzero activations because consonance always increases. To ensure that the networks began in a dissonant state, we initialized some of the unit activations to nonzero values instead of supplying external input. Such assignments of initial activation were done in conformity to procedures used in the relevant psychological experiments and are consistent with the view that dissonance reduction is motivated by an initial imbalance of cognitions.

Network parameters. Because of the large numbers of hand-tuned parameters in constraint satisfaction models, it is important to stress that the present simulations were conducted with a minimum of parameter adjustment. Network weights were excitatory, inhibitory, or zero; resistance was either high or low; and initial levels of activation were either high or low. This is to say that a standard set of parameters sufficed to capture a broad range of dissonance experiments. Furthermore, global randomization techniques revealed that basic effects were quite robust against substantial parameter variation. In general, statistical significance was lost only with randomization of up to 100% of initial parameter values; parameter randomization of up to 50% of initial values always yielded the expected outcomes at conventional levels of statistical significance.

Learning and dissonance. There was no learning of connection weights in our simulations, even though such values can be learned for some constraint satisfaction models (e.g., Ackley,

Hinton, & Sejnowski, 1985; Anderson & Mozer, 1981; Hinton & Sejnowski, 1986). This omission of learning reflects the fact that the typical dissonance experiment is not an occasion for learning. Instead, in a typical dissonance experiment, acculturated and experienced participants commit themselves to some behavior under the influence of a few salient, experimentally engineered cognitions. These cognitions act as constraints on the participant's subsequent reevaluations. In this fashion, the typical dissonance experiment capitalizes on past learning but does not involve new learning.

Indeed, there is a sense in which dissonance reduction processes are antithetical to the process of learning. In experiments on learning, participants are typically able to learn to change their behavior to improve their payoffs, but that learning avenue is closed in dissonance experiments by the fact that participants must remain committed to their behavior. Dissonance reduction is essentially an exercise in coping with behavior that cannot be changed. Perhaps more pointedly, the reduction of cognitive dissonance following decisions may make us less likely to benefit from experience. If each decision we make, after it has been rationalized in the service of dissonance reduction, appears to have been justified, then we will find it particularly difficult to learn from our mistakes. In general, the complex and subtle relations between learning and dissonance have yet to be explored.

Predictions From the Consonance Model

Novel predictions were derived from the consonance network model concerning the locus of reevaluation effects in a free-choice between two relatively undesirable alternatives. Part of the prediction was that most of the separation between evaluations of the chosen and rejected alternatives should be due to an increase in evaluation of the chosen alternative. The other part of the prediction was that the increase in evaluation of the chosen alternative in this difficult choice between two undesirables should be greater than the decrease in evaluation of the rejected alternative in a choice between desirable alternatives. These predictions were confirmed by new psychological research. It is remarkable that the consonance model was able to generate predictions for novel phenomena in a field that has generated so much empirical research over so many years. Indeed, part of the reason that dissonance research has ceased to be so active is that its basic principles and predictions were considered to have been fully explored years ago.

Of some theoretical interest were the plots of dissonance reduction for the various simulations. Although the patterns of dissonance reduction were somewhat different for the different experiments, there were some basic similarities across experiments. First, conditions that should, on dissonance theory grounds, have the most dissonance did in fact start out that way. Second, this high level of dissonance decreased substantially over time cycles. It is unfortunate that there is at present no way of directly assessing dissonance in humans. Until such methods are developed and applied, our dissonance reduction results can be regarded as untested predictions of the consonance constraint satisfaction model.

Extension to Other Dissonance Phenomena

Although the paradigms examined in this article represent the most studied and most reliable derivations from dissonance theory, there are a variety of other phenomena to which dissonance theory had also been applied. These range from the responses of members of doomsday groups to disconfirmations of deeply held beliefs to the transmission of rumors following natural disasters.

A good deal of dissonance research has dealt with the selective exposure paradigm. This concerns the manner in which people seek or avoid additional information that is relevant to a choice they have made. The original prediction was that people would prefer information supporting their choices and avoid information contradicting their choices to reduce dissonance (Festinger, 1957). This paradigm has, in general, generated much more long-term controversy than the insufficient justification and free-choice paradigms because of many results that either failed to support the selective exposure predictions or directly contradicted them by finding a relative preference for dissonant information (cf. review by Freedman & Sears, 1965). However, subsequent research (reviewed in Frey, 1986) inspired by theoretical reformulations (Festinger, 1964) has permitted a more optimistic appraisal of the ability of dissonance theory to deal with selective exposure effects. The theoretical revisions emphasized that consonant information is not always preferred over dissonant information (Festinger, 1964). In particular, there should be a relative preference for dissonant information when this information is perceived to be easily refutable or useful for future decision making. It might be fruitful to see whether consonance networks could capture some of the newer and more replicable selective exposure phenomena.

In recent years, considerable attention has been paid to the study of the arousal and motivational properties of cognitive dissonance (e.g., Cooper & Fazio, 1984; Cooper, Zanna, & Taves, 1978; Zanna & Cooper, 1974). There is also a contemporary focus on the importance of personal responsibility (Cooper & Fazio, 1984) and the self-concept (Steele, 1988; Thibodeau & Aronson, 1992) in the arousal of dissonance. The extent to which such relatively complex social, cognitive, and physiological phenomena can be modeled with constraint satisfaction models remains to be seen.

It would seem that the consonance constraint satisfaction model should, in principle, be able to capture both dissonance arousal and dissonance motivation. Plots of total network dissonance over time cycles in our simulations (Figure 3) indicated that dissonance started at differentially high levels, depending on conditions, and decreased steadily over time, as would be expected. Some psychological experiments found that dissonance arousal could be externally modulated by administration of a drug, such as an amphetamine, tranquilizer, or placebo (Cooper et al., 1978). Such effects might be simulated by introduction of a scalar at the front of Equations 1 and 2. The larger the scalar, the less the overall arousal.

Motivation to reduce dissonance is quite clearly provided by the activation update rules described in Equations 3–5. These equations ensure that dissonance will be reduced, subject to the various constraints supplied by connection weights and initial unit activations. What might be more difficult to simulate in

the arousal literature is a separate intermediate procedure to evaluate dissonance arousal as aversive (Cooper & Fazio, 1984). However, the psychological necessity of having that evaluation as a separate step is perhaps questionable.

Contemporary research on the role of personal responsibility and self-concept might be simulated in consonance constraint satisfaction networks by adding various self-related cognitions and implications. There are no inherent limits on network size in these models, and indeed one of their strengths is the ability to test the dissonance implications of large networks of cognitions. This is not illustrated in our article because we have limited ourselves initially to published dissonance experiments.

Theoretical Unification

Cognitive dissonance phenomena have often been considered as being distinct from less counterintuitive psychological phenomena. Dissonance effects, especially in the insufficient justification paradigms, have been cited as examples that appear to contradict both common sense and other established psychological principles (e.g., Chapanis & Chapanis, 1964; Janis & Gilmore, 1965; Ring, 1967). However, because constraint satisfaction models also account for a wide variety of other phenomena, such as belief revision, explanation, schema completion, analogical reasoning, causal attribution, and content-addressable memory retrieval, their use in this case suggests that cognitive dissonance is not as exotic as it appears. At a deeper theoretical level, that of constraint satisfaction, dissonance is fundamentally related to many other, more mundane psychological processes that can be understood as the progressive application of constraints supplied by personal attitudes, beliefs, and memory traces.

In a similar vein, it is interesting to note that cognitive dissonance theory is but one of a number of theories in social psychology emphasizing that people try to achieve consistency among their cognitions (Abelson et al., 1968; Abelson & Rosenberg, 1958; Heider, 1946, 1958; McGuire, 1960; Newcomb, 1953; Osgood & Tannenbaum, 1955). Although these various cognitive consistency theories have enjoyed considerable success as verbal formulations, the underlying reasoning mechanisms for establishing consistency have never been precisely specified. It is quite likely that connectionist constraint satisfaction models could serve as a general modeling technique and explanatory device in these areas as well (Read & Miller, 1994; Shultz & Lepper, 1992; Spellman & Holyoak, 1992; Spellman et al., 1993). Indeed, we believe that the consonance model presented here could be extended to these other consistency theories.

References

Abelson, R. P. (1968). Psychological implication. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook* (pp. 112–139). Chicago: Rand McNally.

Abelson, R. P., Aronson, E., McGuire, W. J., Newcomb, T. M., Rosenberg, M. J., & Tannenbaum, P. H. (Eds.). (1968). *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally.

Abelson, R. P., & Rosenberg, M. J. (1958). Symbolic psycho-logic: A model of attitudinal cognition. *Behavioral Science, 3,* 1–13.

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science, 9,* 147–169.

Anderson, J. A. (1995). *An introduction to neural networks.* Cambridge, MA: MIT Press.

Anderson, J. A., & Mozer, M. C. (1981). Categorization and selective neurons. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 213–236). Hillsdale, NJ: Erlbaum.

Aronson, E. (1969). The theory of cognitive dissonance: A current perspective. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 4, pp. 1–34). New York: Academic Press.

Aronson, E., & Carlsmith, J. M. (1963). Effect of severity of threat on the devaluation of forbidden behavior. *Journal of Abnormal and Social Psychology, 66,* 584–588.

Aronson, E., & Mills, J. (1959). The effect of severity of initiation on liking for a group. *Journal of Abnormal and Social Psychology, 59,* 177–181.

Brehm, J. W. (1956). Post-decision changes in the desirability of choice alternatives. *Journal of Abnormal and Social Psychology, 52,* 384–389.

Brehm, J. W., & Cohen, A. R. (1962). *Explorations in cognitive dissonance.* New York: Wiley.

Calder, B. J., Ross, M., & Insko, C. A. (1973). Attitude change and attitude attribution: Effects of incentive, choice, and consequences. *Journal of Personality and Social Psychology, 25,* 84–99.

Cartwright, D., & Harary, F. (1956). Structural balance: A generalization of Heider's theory. *Psychological Review, 63,* 277–293.

Chapanis, N. P., & Chapanis, A. C. (1964). Cognitive dissonance: Five years later. *Psychological Bulletin, 61,* 1–22.

Collins, B. E., & Hoyt, M. F. (1972). Personal responsibility-for-consequences: An integration and extension of the "forced compliance" literature. *Journal of Experimental Social Psychology, 8,* 558–593.

Cooper, J., & Fazio, R. H. (1984). A new look at dissonance theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 17, pp. 229–266). New York: Academic Press.

Cooper, J., Zanna, M. P., & Taves, P. A. (1978). Arousal as a necessary condition for attitude change following forced compliance. *Journal of Personality and Social Psychology, 36,* 1101–1106.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association, 50,* 1096–1121.

Feldman, S. (Ed.). (1966). *Cognitive consistency.* New York: Academic Press.

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

Festinger, L. (1964). *Conflict, decision, and dissonance.* Stanford, CA: Stanford University Press.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology, 58,* 203–210.

Freedman, J. L. (1965). Long-term behavioral effects of cognitive dissonance. *Journal of Experimental Social Psychology, 1,* 145–155.

Freedman, J. L., & Sears, D. O. (1965). Selective exposure. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 57–97). New York: Academic Press.

Frey, D. (1986). Recent research on selective exposure to information. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 41–80). New York: Academic Press.

Gerard, H. B., & Mathewson, G. C. (1966). The effects of severity of initiation on liking for a group: A replication. *Journal of Experimental Social Psychology, 2,* 278–287.

Greenwald, A. G., & Ronis, D. L. (1978). Twenty years of cognitive dissonance: Case study of the evolution of a theory. *Psychological Review, 85,* 53–57.

Heider, F. (1946). Attitudes and cognitive organization. *Journal of Personality, 21,* 107–112.

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

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computation.* Reading, MA: Addison-Wesley.

Hinton, G. E., & Sejnowski, T. J. (1986). Learning and relearning in Boltzmann machines. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 282–317). Cambridge, MA: MIT Press.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13,* 295–355.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, USA, 79,* 2554–2558.

Hopfield, J. J. (1984). Neurons with graded responses have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences, USA, 81,* 3008–3092.

Janis, I. L., & Gilmore, J. B. (1965). The influence of incentive conditions on the success of role playing in modifying attitudes. *Journal of Personality and Social Psychology, 1,* 17–27.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review, 95,* 163–182.

Lepper, M. R. (1973). Dissonance, self-perception, and honesty in children. *Journal of Personality and Social Psychology, 25,* 65–74.

Lepper, M. R. (1983). Social-control processes and the internalization of values: An attributional perspective. In E. T. Higgins, D. N. Ruble, & W. W. Hartup (Eds.), *Social cognition and social development* (pp. 294–330). New York: Cambridge University Press.

Linder, D. E., Cooper, J., & Jones, E. E. (1967). Decision freedom as a determinant of the role of incentive magnitude in attitude change. *Journal of Personality and Social Psychology, 6,* 245–254.

McGuire, W. J. (1960). A syllogistic analysis of cognitive relationships. In C. I. Holland & M. J. Rosenberg (Eds.), *Attitude organization and change* (pp. 65–111). New Haven, CT: Yale University Press.

Newcomb, T. M. (1953). An approach to the study of communicative acts. *Psychological Review, 60,* 393–404.

Osgood, C. E., & Tannenbaum, P. H. (1955). The principle of congruity in the prediction of attitude change. *Psychological Review, 62,* 42–55.

Pepitone, A., McCauley, C., & Hammond, P. (1967). Change in attractiveness of forbidden toys as a function of severity of threat. *Journal of Experimental Social Psychology, 3,* 221–229.

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* (pp. 209–235). Hillsdale, NJ: Erlbaum.

Ring, K. (1967). Experimental social psychology: Some sober questions about some frivolous values. *Journal of Experimental Social Psychology, 3,* 113–123.

Rosenberg, M. J., & Abelson, R. P. (1960). An analysis of cognitive balancing. In M. J. Rosenberg, C. I. Hovland, W. J. McGuire, R. P. Abelson, & J. W. Brehm (Eds.), *Attitude organization and change* (pp. 112–163). New Haven, CT: Yale University Press.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. (1986). Schemata and sequential thought processes in PDP models. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 7–57). Cambridge, MA: MIT Press.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures.* Hillsdale, NJ: Erlbaum.

Schneider, W. (1988). Sensitivity analysis in connectionist modeling. *Behavior Research Methods, Instruments, & Computers, 20,* 282–288.

Shultz, T. R. (1992). *Integrating causal heuristics in a constraint satisfaction model.* Technical Report No. 92-5-26, McGill Cognitive Science Centre, McGill University, Montréal.

Shultz, T. R., & Lepper, M. R. (1992). A constraint satisfaction model of cognitive dissonance phenomena. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 462–467). Hillsdale, NJ: Erlbaum.

Shultz, T. R., Léveillé, E., & Lepper, M. R. (1995). *Free-choice and cognitive dissonance revisited.* Forthcoming.

Sloman, S. (1990). *Persistence in memory and judgment: Part–set inhibition and primacy.* Doctoral dissertation, Stanford University.

Spellman, B. A., & Holyoak, K. J. (1992). If Saddam is Hitler, then who is George Bush? Analogical mapping between systems of social roles. *Journal of Personality and Social Psychology, 62,* 913–933.

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,* 147–165.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 261–302). New York: Academic Press.

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

Thibodeau, R., & Aronson, E. (1992). Taking a closer look: Reasserting the role of the self-concept in dissonance theory. *Personality and Social Psychology Bulletin, 18,* 591–602.

Thompson, M. M., Zanna, M. P., & Griffin, D. W. (1995). Let's not be indifferent about (attitude) ambivalence. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 361–386). Hillsdale, NJ: Erlbaum.

Wicklund, R. A., & Brehm, J. W. (1976). *Perspectives on cognitive dissonance.* Hillsdale, NJ: Erlbaum.

Zanna, M. P., & Cooper, J. (1974). Dissonance and the pill: An attribution approach to studying the arousal properties of dissonance. *Journal of Personality and Social Psychology, 29,* 703–709.

Zimbardo, P. G. (1969). *The cognitive control of motivation.* Glendale, IL: Scott, Foresman.

Received November 14, 1994
Revision received August 11, 1995
Accepted August 25, 1995 ■