Simulation of Self-affirmation Phenomena in Cognitive Dissonance

Thomas R. Shultz (shultz@psych.mcgill.ca)
Department of Psychology; McGill University
Montreal, QC H3C 1B1 Canada

Mark R. Lepper (lepper@psych.stanford.edu)
Department of Psychology; Stanford University
Stanford, CA 94305-2130 USA

Abstract

The consonance constraint-satisfaction model, which has simulated the major paradigms of classical cognitive dissonance theory, is here extended to deal with more contemporary findings concerning self-affirmation phenomena in dissonance reduction. The key addition to the model, which has also figured in recent simulations of arousal phenomena, is to lessen activity level within the neural network model in self-affirmation conditions. These and other simulations continue to show that dissonance phenomena can be explained in terms of constraint satisfaction.

Introduction

One of the fundamentally important theories in social psychology is cognitive dissonance theory, which has generated a literature of more than 1000 studies over the past 40 years (Festinger, 1957; Thibodeau & Aronson, 1992). We have recently modeled a number of the central dissonance phenomena using constraint-satisfaction neural networks (Shultz & Lepper, 1996, 1998a&b, 1999a&b). Our so-called consonance model covered insufficient justification, free choice, arousal, and some self-concept phenomena. The model also predicted new free-choice effects that were subsequently confirmed by further psychological experimentation (Shultz, Léveillé, & Lepper, 1999). In this paper, we report on an extension of the model to deal with a prominent self-concept effect in dissonance called self-affirmation.

Dissonance is hypothesized to occur when behavior is inconsistent with self-concept (Steele, 1988; Thibodeau & Aronson, 1992). Because most people have a positive self-concept, behaviors such as lying or trying to persuade others of a position that one does not agree with arouse dissonance and lead to attitude change that reduces the dissonance. However, f important aspects of the self-concept have been recently affirmed, even aspects irrelevant to an experimentally induced inconsistency, there may be no need to reduce dissonance via attitude change. Steele (1988) presented experiments in which fairly subtle self-affirmation manipulations eliminated dissonance effects. Some of these experiments concern insufficient justification via forced compliance, and others deal with free choice. We return to these experiments after reviewing the consonance model used in the simulations.

The Consonance Model

The consonance model holds that dissonance reduction is a constraint satisfaction problem. The motivation to reduce dissonance stems from the various soft constraints on the beliefs and attitudes that an individual holds. A consonance network corresponds to a person's representation of the situation created in the conditions of a dissonance experiment. Activations of network units represent the direction and strength of a person's cognitions. Weights between cognitions represent psychological implications. These unit activations and weights may vary across the different conditions of a single experiment.

Consonance is the degree to which similarly evaluated units are linked by excitatory weights and oppositely valued units are linked by inhibitory weights. More formally, consonance in a network is defined by

$$consonance = \sum_i \sum_j w_{ij} a_i a_j$$

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.

Activation spreads over time cycles by two update rules:

$$a_i(t+1) = a_i(t) + net_i\left(ceiling - a_i(t)\right) \text{ when } net_i \geq 0$$

$$a_i(t+1) = a_i(t) + net_i\left(a_i(t) - floor\right) \text{ when } net_i < 0$$

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 maximum activation, *floor* is the minimum activation, and net_i is the net input to unit i, defined as:

$$net_i = resist_i \sum_j w_{ij} a_j$$

where $resist_i$ refers to the resistance of receiving unit i to having its activation changed.

At each time cycle, n units (normally the number of units in the network) are randomly selected and updated. The update rules ensure that consonance increases or stays the same across cycles. Consonance increases because positive net inputs drive unit activations toward the ceiling and negative net inputs drive them toward the floor. Consonance increases until units reach extreme values or net inputs fall to 0. When consonance reaches asymptote, updating stops.

Consonance networks are hand-built to implement particular dissonance experiments using a set of five principles that map dissonance theory to the consonance model:

1. A cognition is implemented by the net activation of a pair of negatively connected units, one of which represents the positive aspect and the other the negative aspect of the cognition.
2. Cognitions are connected to each other based on their causal implications.
3. Dissonance is the negative of consonance divided by the number of nonzero inter-cognition relations.
4. Networks settle into more stable, less dissonant states as unit activations are updated.
5. Unit activations, but not connection weights, are allowed to change, and some cognitions are more resistant to change than others. In particular, beliefs, behaviors, and justifications are more resistant to change than are evaluations and attitudes.

Additional details about the consonance model and its assumptions are available in our previous papers (Shultz & Lepper, 1996, 1998a).

Forced Compliance

Forced compliance is the most popular dissonance technique within the most prominent dissonance paradigm of insufficient justification. Insufficient justification concerns cases in which a person does something inconsistent with his or her attitudes without much justification. The less the justification, the more cognitive dissonance is created.

In a forced-compliance experiment (Steele, 1988, p. 272), college students were selected for their strong opposition to an increase in tuition fees. They were then persuaded to write essays supporting a substantial tuition increase. In one condition, they were given a choice about whether to write the essay; in another condition, they were given very little choice about whether to write the essay. When a person freely agrees to argue against personal beliefs, this creates dissonance, which can be reduced by changing attitudes in the direction of the argument. There should be little or no dissonance when one is pressured to make such arguments.

Before measuring post-experimental attitudes, some participants were first asked to complete the political sub-scale of the Allport-Vernon Study of Values. One-half of them had been previously assessed as having a strong economic-political value orientation, whereas the others did not have this value orientation. Completing the political value scale was supposed to affirm a valued self-concept only for those students with a strong economic-political value orientation.

As shown by the solid line in Figure 1, there was the familiar dissonance effect of more attitude change under high choice than under low choice. Moreover, as predicted, self-affirmation eliminated attitude change, even under high choice conditions. Two other experiments with minor variations yielded similar results (Steele, 1988).

Method

Network specifications for the three conditions are shown in Table 1. There are two relevant cognitions, attitude and essay, and relations between them. As in our previous simulations, each cognition is implemented with a pair of negatively related units, one to represent the positive aspect of the cognition and the other to represent the negative aspect. Net activation for a cognition is computed as activation on the positive unit minus activation on the negative unit. Positive relations between cognitions are implemented by positive weights between their positive units and between their negative units, and negative weights between the positive unit of one cognition and the negative unit of the other cognition. All weights are bi-directional.

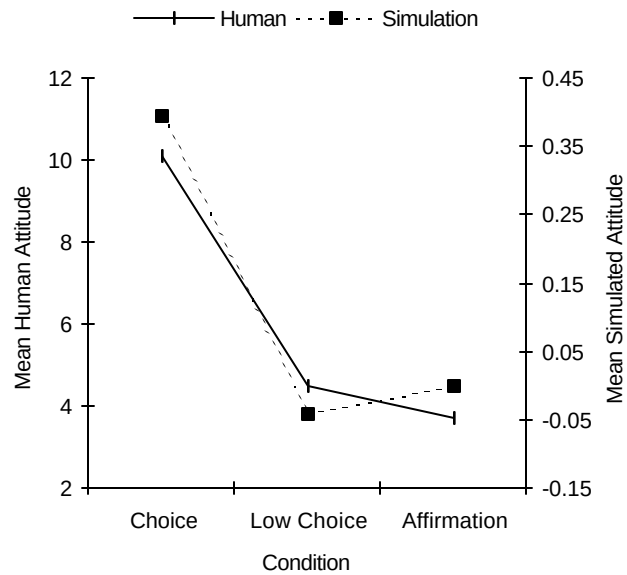

Figure 1: Mean attitude following forced compliance.

All weights and initial unit activations are assigned either high (0.5) or low (0.1) values, according to the five mapping principles described earlier and the descriptions of the experiments being modeled. The floor parameter is 0; the ceiling parameter for positive units is set to 1, and that for negative units is set to 0.5. A *cap* parameter is set to -0.5. This corresponds to the value of the weight between each unit and itself and it prevents activations from growing to ceiling. The *resist* parameter is set to 0.5 for low resistance, and 0.01 for high resistance. These parameter settings are standard across all our dissonance simulations, and some justification for them is provided in our longer papers, (Shultz & Lepper, 1996, 1998a, 1999a).

Table 1: Network specifications for forced compliance.

Condition	Attitude	Essay	Relation
Choice	-0.5	0.5	0.5
Low Choice	-0.5	0.5	0.1
Affirmation	-0.25	0.25	0.25

In this experiment, there is a positive relation between attitude and essay because the more positive one's attitude toward tuition increases, the more likely one would be to

agree to write an essay in favor of tuition increases. This relation is high in the choice condition and low in the low-choice condition. Initially, attitude is given a high negative value to reflect students' initial attitudes; and essay is given a high positive value because the essay was indeed written by all students. An activity-level scalar of 0.5 (the same value used in our other simulations of arousal and self concept) reduces initial activations and weights in the self-affirmation condition, relative to the no-affirmation conditions. The theoretical justification for using a scalar in this way is that self-affirmation is hypothesized to reduce the importance of a dissonant situation (Steele, 1988, p. 292).

All initial unit activations and weights are randomized for each network by adding or subtracting a random proportion of their initial amounts. The three proportion ranges in which additions or subtractions are randomly selected under a uniform distribution are .1, .5, and 1. This increases psychological realism because not everyone can be expected to share the same parameter values. It also allows a test of robustness of the model. Twenty networks were run in each condition at these three different levels of parameter randomization. Networks were run for 30 cycles, which was sufficient to approach asymptotic activation levels.

Results

Mean attitude toward the view espoused in the essay is presented, in the dashed line in Figure 1, for networks at the .5 level of parameter randomization. As with Steele's (1988) subjects, attitudes are more positive under choice than under the other two conditions. An ANOVA with condition as the single factor revealed significant main effects of condition, $F(2, 57) = 67$, $p < .001$. A contrast F with weights of +2 for choice, -1 for low choice, and -1 for self-affirmation is significant $F(1, 57) = 135$, $p < .001$, with no significant residual, $F(1, 57) < 1$. Proportion of total variance accounted for by this F is .99.

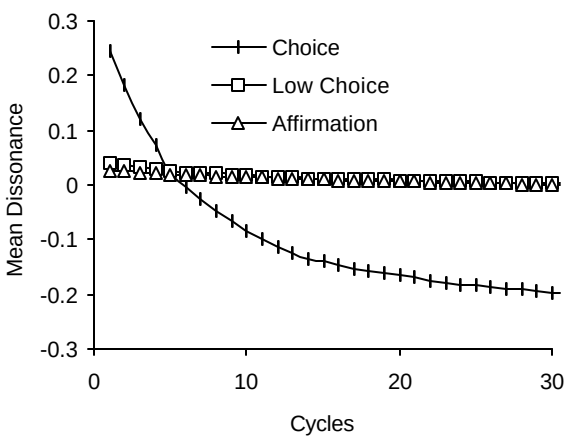

Figure 2: Mean dissonance following forced-compliance.

Mean dissonance scores over time cycles, for networks run at .5 parameter randomization for the three conditions, are shown in Figure 2. Dissonance starts high in the choice condition and is greatly reduced over time. In contrast, there is minimal dissonance in the other two conditions and very little dissonance reduction. Similar results were obtained at parameter randomization levels of .1 and 1.

Discussion

The consonance networks provide a good fit to the attitude change data reported by Steele (1988). There is considerable attitude change in the choice condition, but very little in the low-choice and self-affirmation conditions. There is also a close correspondence between amount of attitude change and plots of dissonance reduction in that the condition with sharp dissonance reduction is also the one with the most attitude change. Examination of dissonance plots is a bonus of computer simulations -- there is no known way to measure dissonance directly in humans. Such plots of simulated dissonance can help to understand the more indirect attitude-change effects that occur as a way of reducing dissonance.

Free Choice

Steele (1988, p. 276) also presents a free-choice experiment that shows self-affirmation effects. Participants rated and ranked 10 music albums and were then given a choice to keep either their fifth- or sixth-ranked album. Choosing between qualitatively distinct objects creates dissonance because the chosen object is less than perfect and the rejected object has some desirable features that are forgone when an irreversible choice is made. The dissonance arising from a free choice is typically reduced by increasing evaluation of the chosen object and decreasing evaluation of the rejected object (Brehm, 1956; Shultz et al., 1999).

In Steele's experiment, one-half of the participants had been previously selected for having a strong scientific-value orientation and for indicating that a lab coat symbolized these values. The others did not share these values. One-half of the participants in each of these groups were asked to wear a lab coat for the rest of the experiment, during which they rated the albums again, after making their choices.

Post-decisional spread of alternatives was measured by adding the increase in the value of the chosen item and the decrease in the value of the rejected item. There were three control conditions, one with participants not having a science orientation and not wearing a lab coat, another with participants not having a science orientation but wearing a lab coat, and a third with participants having a science orientation but not wearing a lab coat. There were identical dissonance effects in all three control conditions, but not for the self-affirmed, scientifically-oriented students wearing a lab coat. Mean spread of alternatives was higher in the control conditions than in the self-affirmation condition, as shown by the solid line in Figure 3. Once again, apparently irrelevant self-affirmation precluded dissonance reduction.

Method

Network specifications for these two groups of conditions are shown in Tables 2 and 3. There are three cognitions: a decision and evaluations of the chosen and the rejected objects. Because the decision is public and irreversible, it has

high resistance and high initial activation; the two evaluations have low resistance. Initial evaluation of the chosen object is somewhat higher than that for the rejected object because people generally choose items that they rate higher.

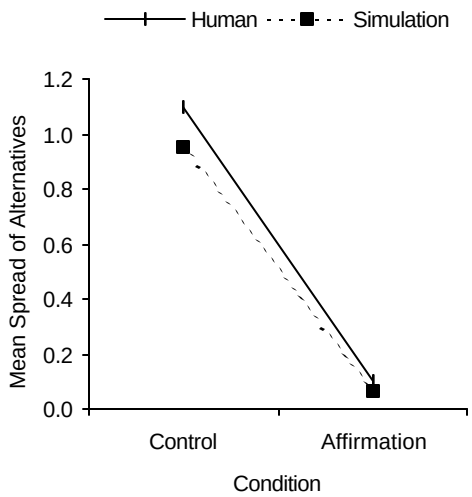

Figure 3: Mean spread of alternatives following free choice.

The relation between the decision and the chosen object is positive because the better-liked object is chosen. The two objects are negatively related because they compete for an exclusive choice. Both relations have high values in the control condition. To implement self-affirmation, initial activations and weights are scaled by .5. Networks in each condition were run for 40 cycles, which was sufficient for saturation. As is customary in our simulations, all weights and initial unit activations were randomized at up to .1, .5, or 1 of the values shown in Tables 2 and 3. Other parameter settings are also the same as in our other dissonance simulations.

Table 2: Initial net activations for free choice.

	Condition	
Cognition	Control	Affirmation
Chosen	.30	.15
Rejected	.20	.10
Decision	.50	.25

Results

Spread between evaluations of the two choices was computed as in Steele (1988). Change in evaluation of each object is the difference between initial evaluation and evaluation after 40 cycles. Spreading of alternatives is the sum of the increase in evaluation of the chosen alternative and the decrease in evaluation of the rejected alternative. Mean spreading of the alternatives is plotted, on the dashed line in Figure 3, at the .5 level of parameter randomization. There is a larger spread of the alternatives in the control than in the self-affirmation condition, $F(1, 38) = 76$, $p < .001$.

Mean dissonance scores across time cycles in networks at .5 parameter randomization are shown in Figure 4 for the two conditions. Although dissonance starts low in both conditions, it drops only in the control condition. Similar results were found at parameter randomizations of .1 and 1.

Table 3: Relations between cognitions for free choice.

Relation of	Condition	
chosen to	Control	Affirmation
Decision	.50	.25
Rejected	-.50	-.25

Discussion

Consonance networks yield greater separation of alternatives in the control than in the self-affirmation condition, as found with human participants (Steele, 1988). Dissonance reduction is also greater in the control than in the self-affirmation condition, consistent with the idea that attitude change is driven by dissonance reduction.

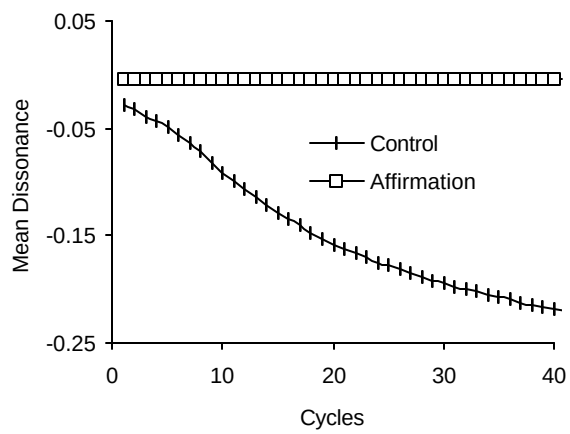

Figure 4: Mean dissonance following free choice.

General Discussion

These simulations extend the consonance model to rather subtle aspects of dissonance reduction involving the self-concept, using the same conventions, mapping principles, and default parameter values as in previous simulations. In all of these cases, dissonance arises when constraints between simultaneously held cognitions are unsatisfied. Dissonance is reduced as the constraints are satisfied, typically by changing evaluations of entities in the situation defined by the dissonance experiment. The self-affirmation phenomena considered here had not previously been simulated and were not generally seen as being closely related to other contemporary dissonance phenomena on emotional arousal. As in earlier simulations, the consonance model is here shown to be robust against parameter variation, as revealed by the fact that even a high degree of parameter randomization does not affect the pattern of overall results.

A key, unifying concept in simulating contemporary dissonance phenomena in self-concept and arousal is that of activity level. An activity scalar adjusts the overall level of activation in networks that represent dissonant situations. In the present simulations, the activity-level scalar operates

much like a tranquilizing drug in arousal simulations (Shultz & Lepper, 1999b), by decreasing activation of the representation of the dissonant situation.

Self-affirmation manipulations are thus hypothesized to decrease the relative importance of being in a dissonant situation. When you feel good about yourself, being in a dissonant situation is not nearly so bothersome, and you become immune to the effects of dissonance reduction. This reveals a somewhat unexpected theoretical communality between arousal and self-concept effects.

This analysis is consistent with recent results on *trivialization* as a mode of dissonance reduction (Simon, Greenberg, & Brehm, 1995). Merely making salient to participants asked to write counter-attitudinal essays the contrast between issues they believe to be of great consequence and the less important topic of their own essays reduces attitude change in the direction of the position advocated.

At the level of the brain or an artificial neural network, the key theoretical notion is that of activity level. Dissonance effects are enhanced by increases in activity level and dampened by decreases in activity level. There are a variety of ways to modulate activity level, including general manipulations such as drugs (Cooper, Zanna, & Taves, 1978) and specific manipulations such as attention to particular cognitions (Read & Miller, 1998a). Consequently, activity level has the potential to unify theoretical understanding of several apparently different dissonance phenomena.

The general success of the consonance model enables a theoretical reinterpretation of dissonance that stresses commonalties with other psychological phenomena that result from constraint satisfaction. Phenomena such as analogical reasoning, person perception, schema completion, attitude change, and dissonance reduction can all be understood in terms of the dynamics of constraint satisfaction (Holyoak & Thagard, 1989; Read & Miller, 1998a, b; Rumelhart, Smolensky, McClelland, & Hinton, 1986; Spellman & Holyoak, 1992; Spellman, Ullman, & Holyoak, 1993; Thagard, 1989).

Acknowledgments

This research was supported by a grant to the first author from the Social Sciences and Humanities Research Council of Canada and by grant MH-44321 to the second author from the U.S. National Institute of Mental Health.

References

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

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.

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

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

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

Read, S. J., & Miller, L. C. (Eds.). (1998b). *Connectionist models of social reasoning and social behavior.* Hillsdale, NJ: Erlbaum.

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.

Shultz, T. R., & Lepper, M. R. (1996). Constraint satisfaction modeling of cognitive dissonance phenomena. *Psychological Review, 103,* 219-240.

Shultz, T. R., & Lepper, M. R. (1998a). The consonance model of dissonance reduction. In S. J. Read & L. C. Miller (Eds.), *Connectionist models of social reasoning and social behavior* (pp. 211-244). Hillsdale, NJ: Erlbaum.

Shultz, T. R., & Lepper, M. R. (1998b). A constraint-satisfaction model of Machiavellianism effects in cognitive dissonance. *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 957-962). Hillsdale, NJ: Erlbaum.

Shultz, T. R., & Lepper, M. R. (1999a). Computer simulation of cognitive dissonance reduction. In E. Harmon-Jones & Mills, J. (Eds.), *Cognitive dissonance: Progress on a pivotal theory in social psychology* (pp. 235-265). Washington, DC: American Psychological Association.

Shultz, T. R., & Lepper, M. R. (1999b). Consonance network simulations of arousal phenomena in cognitive dissonance. *Proceedings of the Twenty-first Annual Conference of the Cognitive Science Society* (pp. 659-664). Hillsdale, NJ: Erlbaum.

Shultz, T. R., Léveillé, E., & Lepper, M. R. (1999). Free choice and cognitive dissonance revisited: Choosing "lesser evils" vs. "greater goods." *Personality and Social Psychology Bulletin, 25,* 40-48.

Simon, L., Greenberg, J., & Brehm, J. (1995). Trivialization: The forgotten mode of dissonance reduction. *Journal of Personality and Social Psychology, 68,* 247-260.

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.