A Constraint-Satisfaction Model of Machiavellianism Effects in Cognitive Dissonance

Thomas R. Shultz (shultz@psych.mcgill.ca)
Department of Psychology; McGill University; 1205 Penfield Avenue
Montreal, Quebec H3A 1B1 Canada

Mark R. Lepper (lepper@psych.stanford.edu)
Department of Psychology; Stanford University; Jordan Hall, Building 420
Stanford, CA 94305-2130 USA

Abstract

The consonance constraint-satisfaction model is applied to Machiavellianism self-concept effects in cognitive dissonance. Networks parameterized for low Machiavellian traits showed the usual dissonance effect, i. e., more attitude change after giving a counter-attitudinal speech than after not giving such a speech, whereas networks parameterized for high Machiavellian traits showed the reverse, thus capturing human data. Classical dissonance theory had not accounted for the fact that people with high Machiavellian traits showed less attitude change after giving a counter-attitudinal speech than after not giving such a speech. The model predicts initial dissonance and the course of dissonance reduction in the various experimental conditions. The results underscore the point that cognitive dissonance operates according to the same constraint-satisfaction principles that govern a variety of other psychological phenomena.

Introduction

If you are induced to publicly support a position with which you initially disagree, and are provided with rather little justification for doing so, it is very likely that your attitude will change in the direction of your argument. Somewhat paradoxically, the less justification that you receive for making the public argument, the more your attitude will change in that direction. Such compelling, but somewhat counter-intuitive, results have long been the hallmark of a psychological theory known as cognitive dissonance (Festinger, 1957).

Briefly, cognitive dissonance theory holds that we experience dissonance among our attitudes and beliefs as psychological discomfort, and that we try to reduce this dissonance by changing our attitudes in order to increase consistency as much as we can without creating new dissonance. For the last four decades, cognitive dissonance theory has been the subject of over 1000 publications and has become a central foundation of social psychology. However, the underlying cognitive mechanisms for dissonance phenomena have not, until very lately, been adequately specified, and the theory has traditionally been viewed as distinct from those concerning more mundane psychological processes.

A variety of phenomena from the major paradigms of cognitive dissonance theory, known in the psychological literature as insufficient justification and free-choice, have recently been modeled with constraint-satisfaction neural networks (Shultz & Lepper, 1992, 1996). In several cases, phenomena were covered more accurately by this so-called consonance model than they were by classical dissonance theory. Superior coverage was due to the inclusion of constraints not present in dissonance theory and to the increased precision inherent to the computational formulation. Some predictions of the consonance model have been confirmed by new psychological research on free-choice (Shultz, Leveille, & Lepper, in press).

The success of the consonance model enables a reinterpretation of cognitive dissonance and its reduction that underscores commonalties with several other psychological phenomena that appear to be governed by constraint-satisfaction, including content-addressable memory, comprehension, revision of beliefs, explanation, attitude change, and person perception (Holyoak & Thagard, 1989; Kintsch, 1988; Kunda & Thagard, 1996; Read & Miller, 1994, 1998; Rumelhart, Smolensky, McClelland, & Hinton, 1986; Sloman, 1990; Spellman & Holyoak, 1992; Spellman, Ullman, & Holyoak, 1993; Thagard, 1989).

This paper reports new simulations of phenomena in the cognitive dissonance literature concerning the role of the self-concept in the experience of dissonance. People who agree with Machiavellian ideas have been shown to experience and reduce dissonance in a different manner than other individuals (Epstein, 1969). Because Machiavellians do not adhere to traditional moral prohibitions against manipulative duplicity, they apparently do not experience dissonance from arguing for a view that is contradictory to their own. We start with an overview of the consonance model, proceed to a review of psychological results on dissonance and Machiavellianism, and then present the simulations.

The Consonance Model

The consonance model is based on the idea that dissonance reduction can be interpreted in terms of constraint-satisfaction. Dissonance reduction and other modes of consistency seeking (Abelson, Aronson, McGuire, Newcomb, Rosenberg, & Tannenbaum, 1968) can be solved by the satisfaction of numerous soft constraints that can vary in their relative importance. Soft constraints are those that are desirable, but not essential, to satisfy.

Consonance networks can be used to represent an individual's interpretation of a situation created by the experimental setting of a cognitive dissonance experiment. Activations on units in the network represent the strength and direction of the person's attitudes and beliefs. Units can 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 the psychological implications among attitudes and beliefs. These connections between units can be excitatory, inhibitory, or nonexistent. Both unit activations and connection weights can vary across the different conditions of an experiment. Consonance is roughly the degree to which similarly evaluated units are linked by excitatory weights and oppositely valued units are linked by inhibitory weights. Activations change over time cycles to increase consonance while satisfying the various constraints introduced by initial activations and connection weights.

More precisely, the consonance contributed by a particular unit i is

$$consonance_i = \sum_j w_{ij} a_i a_j \quad (1)$$

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. Consonance over the entire network is the sum of the values produced by Equation 1 over all receiving units in the network

$$consonance_n = \sum_i \sum_j w_{ij} a_i a_j \quad (2)$$

Activation spreads across the network over time cycles following these two update rules:

$$a_i(t+1) = a_i(t) + net_i(ceiling - a_i(t)), \text{ when } net_i \geq 0 \quad (3)$$

$$a_i(t+1) = a_i(t) + net_i(a_i(t) - floor), \text{ when } net_i < 0 \quad (4)$$

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

$$net_i = resist_i \sum_j w_{ij} a_j \quad (5)$$

The parameter $resist_i$ indexes the resistance of receiving unit i to having its activation changed.

At each time cycle, n units are randomly selected and updated via Equations 3-5. Typically n is the number of units in the network. The updates produced by Equations 3-5 ensure that consonance either increases or remains constant across cycles. When consonance reaches asymptote, the updating process is stopped.

A generic consonance network can be used to instantiate any particular dissonance experiment consistent with 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 pole and the other represents the negative pole of the cognition. Net activation for the cognition is the difference between the activation on the positive unit and the activation on the negative unit. By default, the activation floor is 0; the ceiling is 1 for positive poles and 0.5 for negative poles.

2. Cognitions are connected to each other based on their assumed causal implications. A negative implication is represented by a negative relation between two cognitions and a positive implication is represented by a positive relation between the cognitions. Connection weights range between -1 and 1, with 0 representing a lack of causal relation. The connection scheme for two generic, positively related cognitions is illustrated in Figure 1. When two cognitions are positively related, 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. These links are reversed for cognitions that are negatively related (not shown in the Figure). Each unit has an inhibitory self-connection specified by the *cap* parameter, described later. All connection weights are bi-directional (also not shown in the Figure). Connection weights have an initial default value of 0.5.

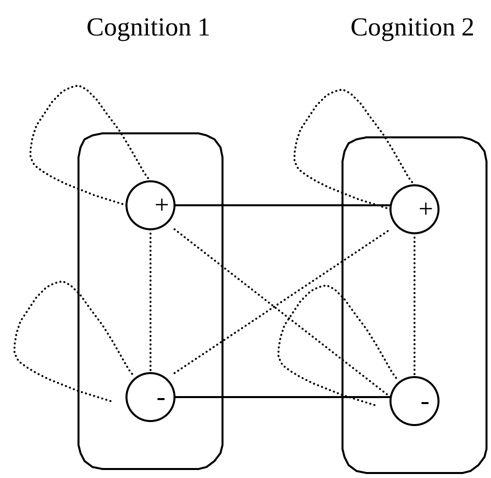

Figure 1: Connection scheme for two cognitions that are positively related. Each cognition, enclosed in a rounded rectangle, has two poles, one positive and the other negative. Excitatory connection weights are indicated by solid lines and inhibitory connection weights by dashed lines.

3. Total dissonance is defined as the negative of total consonance divided by r, the number of inter-cognition relations that are not zero

$$dissonance = \frac{-\sum_i \sum_j w_{ij} a_i a_j}{r} \qquad (6)$$

Dividing by r standardizes the dissonance index across different networks by controlling for the number of relevant relations. Self-connections, designated by w_{ii}, are excluded from this computation of dissonance. This definition of dissonance differs from Festinger's (1957) in that it is formalized, indexes the amount of dissonance in each inter-cognition relation, includes within-cognition ambivalence and consonant relations, and may vary even when all relations are dissonant or all are consonant.

4. Networks tend to settle into more stable, less dissonant states as unit activations are updated with Equations 3-5. This settling is influenced by two parameters. A *cap* parameter with a default of -0.5, corresponding to the value of the connection between each unit and itself, w_{ii}, prevents activations from reaching their ceiling. Weights, resistances, caps, and initial activations are randomized by adding or subtracting a random proportion of their initial amounts. The *rand%* parameter specifies the range, in terms of a proportion, in which additions or subtractions are randomly selected under a uniform distribution. The main purpose of this randomization of parameters is to efficiently assess the robustness of the simulation across variations in parameters. It also increases psychological realism because not everyone can be assumed to share precisely the same parameter values. Further, it violates a connection weight symmetry assumption such that w_{ij} - w_{ji} and thus decreases the stability of network solutions. Typical *rand%* values of 0.1 (low), 0.5 (medium), and 1.0 (high) are used (Shultz & Lepper, 1996).

5. Cognition unit activations, but not connection weights, are allowed to change, and some cognitions are more resistant to change than others, as implemented in Equation 5. In particular, beliefs and justifications are more resistant to change than are evaluations. The *resist* parameter has default values of 0.5 (low) or 0.01 (high). As shown in Equation 5, the larger the resistance parameter, the more readily a unit changes its activation. Additional details about the consonance model and support for its various assumptions are presented in Shultz and Lepper (1996).

Machiavellianism, Self-concept, and Dissonance

Relatively recently, there has been concern with the role of the self-concept in the arousal and reduction of dissonance (Steele, 1988; Thibodeau & Aronson, 1992). The basic idea is that dissonance occurs because a person's behavior is inconsistent with his or her self-concept. Because most people possess a relatively positive self-concept, behaviors such as lying or supporting a disagreeable position arouse dissonance. However, people possessing a self-concept that allows duplicitous manipulation of others would not likely experience dissonance from performing such behaviors.

Initial support for the importance of the self-concept in dissonance came from studies of the trait of Machiavellianism. People scoring low on Machiavellianism seem to experience dissonance in a situation of insufficient justification via forced compliance, but those scoring high on this trait do not (Epstein, 1969). In Epstein's experiment, people were induced, with rather little justification, to write an essay that was contrary to their own attitudes. For those high on the trait of Machiavellianism, lying or writing a counter-attitudinal essay would not be inconsistent with their self-concept, and thus they should not experience the dissonance that others would experience. Epstein's undergraduate participants initially supported fluoridation of water, but were persuaded to give a speech against it. They read some anti-fluoridation arguments and were paid a mere $2 to give the speech. In another condition, participants read the same arguments against fluoridation, and gave no speech, but were paid $2 anyway.

Epstein's results are shown in Figure 2. In support of dissonance theory predictions, mean opinion change towards anti-fluoridation was higher for low Machiavellians who gave the speech than for low Machiavellians who did not give the speech. High Machiavellians showed the opposite trend, which Epstein explained by citing evidence that high Machiavellians are more susceptible to persuasion by factual arguments than are low Machiavellians (Harris, 1966). In this case, the factual arguments were presented in the anti-fluoridation readings. Strictly speaking, dissonance theory only explains the results for low Machiavellians.

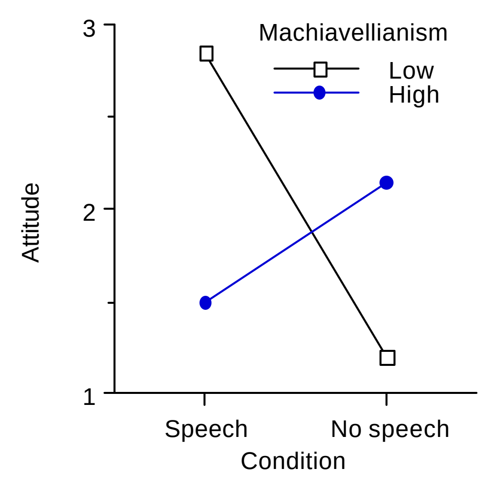

Figure 2: Mean attitude towards the view espoused in a speech (from Epstein, 1969).

Simulations

Network specifications for the four cells of Epstein's (1969) experiment were created following the foregoing five principles that map dissonance theory to constraint-satisfaction models. The specifications are identical to those used in the consonance model simulations (Shultz & Lepper, 1996) of the Linder, Cooper, and Jones (1967) forced compliance experiment with the required exceptions that reflect differences between that experiment and Epstein's (1969).

Specifications for the speech conditions are shown in Figure 3. There are four relevant cognitions, concerning giving the speech, attitude towards the view expressed in the speech, the anti-fluoridation arguments, and the payment. Relations among these cognitions reflect assumed causal relations. The relation between attitude and speech is positive because the more favorable one's attitude, the more likely one would be to agree to make the speech. This relation between attitude and speech is low (0.1) for high Machiavellians, because, for them, there is not a strong relation between attitude and public statement, and high (0.5) for low Machiavellians, for whom there is a strong relation between attitude and public statement. The relation between payment and speech is high and positive (0.5) because a greater payment should increase the likelihood of giving the speech. The relation between attitude and payment is negative (-0.5) because, for any given level of counter attitudinal speech, the more favorable one's attitude, the less one would need to be paid to give the speech. The relation between the anti-fluoridation arguments and anti-fluoridation attitude is positive, but more so for high Machiavellians (0.5) who are supposed to be more influenced by factual arguments than are low Machiavellians (0.1) (Harris, 1966).

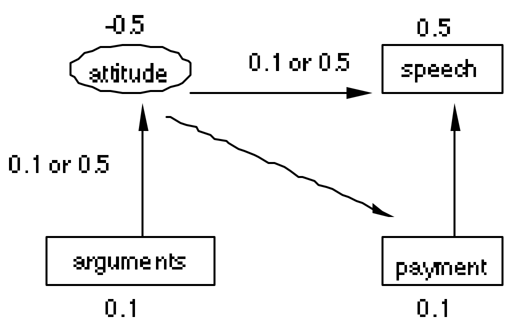

Figure 3: Network design for the speech conditions. In this and Figure 4, cognitions with low resistance are drawn with an ellipse, and those with high resistance are drawn with a rectangle. Positive relations are portrayed with solid arrows and negative relations with dashed arrows. Directions of the arrows reflect the direction of causal influence.

Initial attitude was negative (-0.5) to reflect the actual views of Epstein's (1969) participants. The speech was given an initial positive value (0.5) because it was in fact given by all participants in the speech conditions. Arguments and payment were given low initial values (0.1) to reflect their assumed potency. Attitude, being an evaluation, was given low resistance to change, and the other three cognitions, being beliefs or justifications, were given high resistance to change.

Network specifications for the no speech conditions are shown in Figure 4. Here, there are only three cognitions because the speech was not given. The positive relations between speech and attitude and between speech and pay become 0 under no speech, as there is no speech. The relation between payment and attitude is changed to positive because, with neither of these cognitions connected to giving a speech, receiving a payment might improve one's attitude to the entire experimental session.

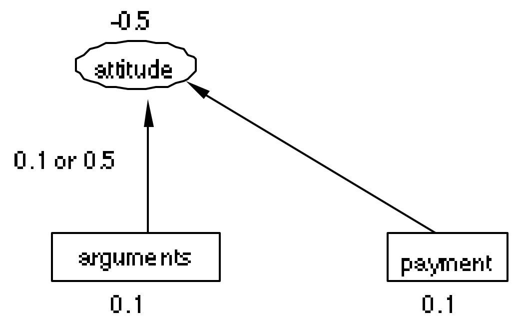

Figure 4: Network design for the no speech conditions.

Twenty networks were run in each of the four conditions at each of the three levels of parameter randomization. Updating continued for 30 cycles, by which time activation asymptotes were reached.

Mean attitude towards the view expressed in the speech is presented in Figure 5 at the highest level of parameter randomization (1.0). Attitude is computed as net activation of the two units representing the attitude cognition, as specified in mapping principle 1. Networks at all three levels of randomization show the cross-over interaction that had occurred with Epstein's human participants. Low Machiavellian networks show the predicted dissonance effect, i.e., more attitude change with a speech than without a speech, while high Machiavellian networks show the reverse. At each level of parameter randomization, there is a statistically reliable Machiavellianism x speech interaction, $p < .001$, $F(1, 76) = 913$ for low randomization, 26.96 for medium randomization, and 12.83 for high randomization.

Although there is no direct measure of dissonance in human participants, the amounts of dissonance experienced by the networks, as computed in Equation 6, can be examined for predictive purposes. Dissonance reduction over time is shown in Figure 6 for the four conditions of the experiment for networks run at high parameter randomization. The low Machiavellianism with speech condition shows the highest initial dissonance and the most dissonance reduction. The other conditions show little initial dissonance and little reduction of dissonance, all of which corresponds to the attitude change results shown in Figure 5.

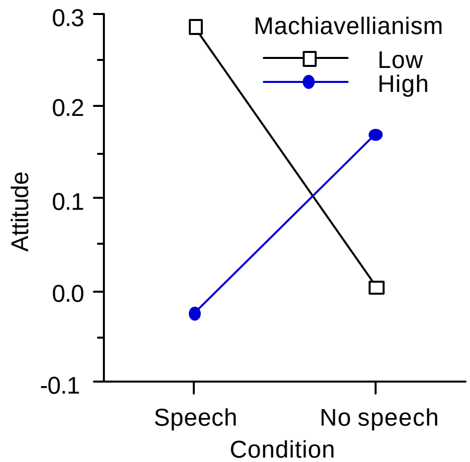

Figure 5: Mean attitude towards the view espoused in a speech, from simulations at a high level of parameter randomization (1.0).

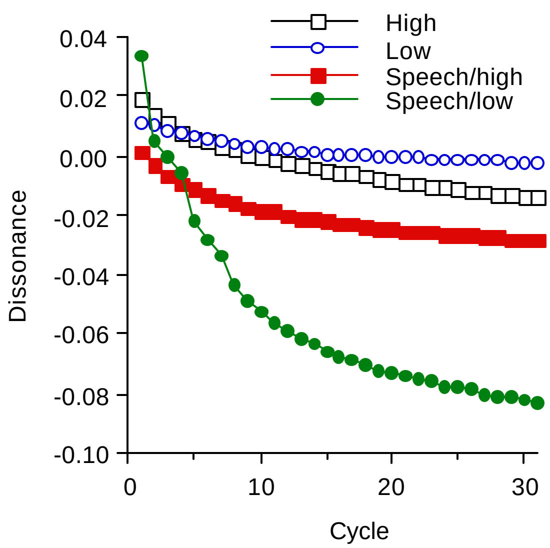

Figure 6: Mean dissonance over time cycles in networks at a high level of parameter randomization (1.0).

Discussion

The simulation results on attitude change mirror those of the human participants in Epstein's (1969) experiment. There is the expected dissonance effect for low Machiavellians and the opposite effect for high Machiavellians. As noted earlier, the results for high Machiavellians are not explainable by classical dissonance theory. As in earlier simulations (Shultz & Lepper, 1996), the inclusion of constraints not present in dissonance theory allows the networks to more closely approximate human data. In this case, the extra constraints concern the relation between attitudes and public statements (generally high, but low for Machiavellians) and the relation of arguments to attitudes (higher for Machiavellians than for others).

The plots of dissonance reduction reveal something about the underlying dynamics in these networks. It is clear from these dissonance plots and those of attitude change that there is a close correspondence between amount of attitude change and amount of dissonance reduction. The higher the initial dissonance, the steeper the dissonance reduction, and the more the change in attitude.

There are a number of other self-concept phenomena in cognitive dissonance that also may be possible to simulate with the consonance model. These phenomena concern the finding that, after affirming an important aspect of the self-concept that may be irrelevant to an experimentally induced inconsistency, people do not need to reduce dissonance by attitude change (Steele, 1988). Such effects have been found in both the free-choice and insufficient justification paradigms of dissonance research.

The fact that reduction of cognitive dissonance can be modeled by constraint-satisfaction networks suggests that dissonance is governed by processes common to a wide variety of other cognitive phenomena (Holyoak & Thagard, 1989; Kintsch, 1988; Kunda & Thagard, 1996; Read & Miller, 1994, 1998; Rumelhart et al., 1986; Sloman, 1990; Spellman & Holyoak, 1992; Spellman et al., 1993; Thagard, 1989). Processes as diverse as memory, belief revision, explanation, attitude formation, person perception, and dissonance reduction all appear to reflect the progressive application of constraints supplied by cognitions and relations between cognitions. There is considerable scope here for theoretical integration and unification using constraint-satisfaction ideas.

Acknowledgments

This research was supported in part by a grant from the Social Sciences and Humanities Research Council of Canada to Shultz and by grant MH-44321 to Lepper from the U.S. National Institute of Mental Health. We are grateful to David Buckingham and Sylvain Sirois for helpful comments on a previous draft.

References

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.

Epstein, G. F. (1969). Machiavelli and the devil's advocate. *Journal of Personality and Social Psychology, 11*, 38-41.

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

Harris, T. M. (1966). *Machiavellianism, judgment independence, and attitude toward teammate in a cooperative group*. Doctoral Dissertation, Columbia University.

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

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

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

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.

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

Read, S. J., & Miller, L. C. (Eds.) (1998). *Connectionistic and PDP 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). Cambridge, MA: MIT Press.

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*. Hillsdale, NJ: Erlbaum.

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

Shultz, T. R., L veill , E., & Lepper, M. R. (in press). Free-choice and cognitive dissonance revisited: Choosing "lesser evils" vs. "greater goods." *Personality and Social Psychology Bulletin*.

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). 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.