Note

1. Address correspondence to Daniel Povinelli, Cognitive Evolution Group, University of Louisiana, 4401 W. Admiral Doyle Dr., New Iberia, LA 70560; e-mail: ceg@louisiana.edu.

References

Call, J., & Tomasello, M. (1999). A nonverbal theory of mind test: The performance of children and apes. *Child Development*, 70, 381–395.

Darwin, C. (1982). *The descent of man*. New York: Modern Library. (Original work published 1871)

de Waal, F.B.M. (1999). Animal behaviour: Cultural primatology comes of age. *Nature*, 399, 635.

Dreifus, C. (1999, August–September). Going ape. *Ms.*, 9(5), 48–54.

Fouts, R. (1997). *Next of kin*. New York: William Morrow and Co.

Goodall, J. (1990). *Through a window*. Boston: Houghton Mifflin.

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, 59, 771–785.

Karin-D'Arcy, R., & Povinelli, D.J. (2002). *Do chimpanzees know what each other see? A closer look*. Manuscript submitted for publication.

Povinelli, D.J. (2000). *Folk physics for apes*. New York: Oxford University Press.

Povinelli, D.J., & O'Neill, D.K. (2000). Do chimpanzees use their gestures to instruct each other? In S. Baron-Cohen, H. Tager-Flusberg, & D.J. Cohen (Eds.), *Understanding other minds* (pp. 459–487). Oxford, England: Oxford University Press.

Povinelli, D.J., Reaux, J.E., Bierschwale, D.T., Allain, A.D., & Simon, B.B. (1997). Exploitation of pointing as a referential gesture in young children, but not adolescent chimpanzees. *Cognitive Development*, 12, 423–461.

Preuss, T.M., & Coleman, G.Q. (in press). Human-specific organization of primary visual cortex: Alternating compartments of dense Cat-301 and calbindin immuno-reactivity in layer 4A. *Cerebral Cortex*.

Russon, A., & Bard, K. (1996). Exploring the minds of the great apes: Issues and controversies. In A.E. Russon, K.A. Bard, & S.T. Parker (Eds.), *Reaching into thought* (pp. 1–20). Cambridge, England: Cambridge University Press.

Savage-Rumbaugh, S., & Lewin, R. (1994). *Kanzi: The ape at the brink of the human mind*. New York: John Wiley & Sons.

Suddendorf, T., & Whiten, A. (2001). Mental evolution and development: Evidence for secondary representation in children, great apes and other animals. *Psychological Bulletin*, 127, 629–650.

Accepting Threatening Information: Self-Affirmation and the Reduction of Defensive Biases

David K. Sherman and Geoffrey L. Cohen[1]

Department of Psychology, University of California, Los Angeles, Los Angeles, California (D.K.S.), and Department of Psychology, Yale University, New Haven, Connecticut (G.L.C.)

Abstract

Why do people resist evidence that challenges the validity of long-held beliefs? And why do they persist in maladaptive behavior even when persuasive information or personal experience recommends change? We argue that such defensive tendencies are driven, in large part, by a fundamental motivation to protect the perceived worth and integrity of the self. Studies of social-political debate, health-risk assessment, and responses to team victory or defeat have shown that people respond to information in a less defensive and more open-minded manner when their self-worth is buttressed by an affirmation of an alternative source of identity. Self-affirmed individuals are more likely to accept information that they would otherwise view as threatening, and subsequently to change their beliefs and even their behavior in a desirable fashion. Defensive biases have an adaptive function for maintaining self-worth, but maladaptive consequences for promoting change and reducing social conflict.

Keywords

self; bias; self-affirmation; attitude change; health risk

Everyday experience confirms that people's judgments are often biased by their beliefs, desires, and vested interests. Political partisans draw opposite conclusions from the same evidence. People who engage in risky health behavior resist even well-reasoned appeals to change their conduct. And the explanations offered by athletes following defeats, or by business executives in response to vanishing profits, frequently prove less illuminating than face-saving. Such judgments promote self-worth, insofar as they protect feelings of adaptiveness and adequacy, or suggest that any discrepancy between one's desires and the outcomes obtained is due to external circumstance rather than internal flaw.

At the same time, however, such defensive responses can cause resistance to change. Indeed, people may allow their beliefs to bias their evaluation of new information, and then use that biased evaluation to derive even further confidence in the validity of their beliefs (Lord, Ross, & Lepper, 1979). Because such responses protect self-worth—shielding people from the conclusion that their beliefs or actions were misguided—we refer to them as *defensive*. To the extent that such defensive biases can be reduced, people may be more open to important but potentially threatening information.

For more than 30 years, social psychologists have tried to understand the origins of defensive biases (see Nisbett & Ross, 1980). Our

own research suggests that such biases arise, in part, from a fundamental motivation to protect the perceived worth and integrity of the self (Steele, 1988). Because the motivation to maintain self-worth can prove powerful, people may resist information that could ultimately improve the quality of their decisions. To the extent that information threatens self-worth, or is presented in a manner that threatens self-worth, people may dismiss, deny, or distort it in a fashion that serves to sustain their personal feelings of adaptiveness and integrity. The cost, of course, is that such information, if studied or acted upon, could ultimately increase their adaptiveness.

Our assertion that defensive biases arise from a motivation to maintain self-integrity is supported by research demonstrating that such biases are attenuated by affirmations of alternative sources of self-worth. Our theorizing builds upon Steele's (1988) self-affirmation theory, which proposes that the desire to maintain the perceived worth and integrity of the self constitutes a fundamental goal of the self-regulatory system. When presented with threatening information, people can satisfy this motivation directly: They can defensively neutralize the information, for example, by interpreting mixed evidence in a manner that supports preexisting beliefs or by dismissing attitude-disconfirming evidence; if the threatening information is an undesirable outcome (e.g., business profits are down), they can attribute that outcome to external circumstance rather than internal weakness. Alternatively, people can satisfy the self-affirmation motivation indirectly: They can draw upon alternative sources of self-integrity, for example, by reflecting upon an important value irrelevant to the threat, or by achieving success in an unrelated domain of self-worth.

REDUCING BIASED INTERPRETATIONS OF AMBIGUOUS INFORMATION

When people face ambiguous evidence, they tend to interpret it in a way that favors preexisting beliefs and desires (Kunda, 1987; Lord et al., 1979). For example, political-party loyalists process presidential debates in a biased manner, leaving the debate with increased confidence in their candidate's virtues. Such bias is motivated, in part, by a desire to protect self-worth. To acknowledge the strengths of the "other side" would presumably threaten one's self-image as an informed and rational voter. Ironically, however, the motivation to protect this self-image may produce irrationality: In the face of evidence that neutral parties would characterize as "mixed" at best, partisans may see confirmation of the validity of their beliefs, and embrace their attitudes with even greater conviction (cf. Lord et al., 1979).

We examined the biased interpretation of ambiguous information in the context of the abortion debate (Cohen, Aronson, & Steele, 2000). Pro-choice partisans and pro-life partisans were presented with a debate between two activists on opposite sides of the abortion dispute. Participants who did not receive a self-affirmation showed a confirmation bias. They judged the activist who shared their views on abortion more favorably than the activist who did not. However, this confirmation bias was sharply attenuated among participants who affirmed a valued source of self-worth (by writing about a personally important value, such as their relations with friends). Self-affirmed individuals rated the activist who shared their views less positively than did their nonaffirmed peers. Moreover, although all participants left the debate feeling more confident in their beliefs about abortion than they had before, this polarization in attitude was significantly reduced among self-affirmed participants (cf. Lord et al., 1979). Note that the affirmation in this study ameliorated bias not in evaluations of evidence but in judgments of disputants—a result particularly germane to mediators who want to encourage trust between parties in conflict.

Bias in the interpretation of ambiguous information can also prevent adaptive changes in health behavior. People for whom a health message is of high personal relevance are more likely to scrutinize that information for fault than are people for whom the message has no special relevance (Kunda, 1987). Thus, individuals who have the most to gain from health communications are often the least likely to accept them. For example, studies have found that coffee drinkers are more critical than non-coffee drinkers of scientific evidence linking caffeine to breast cancer, although they readily accept evidence disconfirming that link (Kunda, 1987).

We examined biased interpretation of ambiguous health information in the context of breast cancer prevention (Sherman, Nelson, & Steele, 2000). Participants were women who were either coffee drinkers or non-coffee drinkers, and they reviewed a scientific report linking caffeine consumption to breast cancer.[2] The reported evidence derived from studies that, although persuasive in some respects, contained significant methodological weaknesses. As in earlier research (e.g., Kunda, 1987), coffee drinkers were more critical of those weaknesses and thus more resistant to the message as a whole than were non-coffee drinkers. Yet, among self-affirmed participants, who had reflected upon a personally important value, coffee drinkers were more open to the information contained in the report than

were non-coffee drinkers, and they intended to reduce their coffee drinking accordingly. When people who would otherwise feel threatened by a health message affirmed an alternative source of self-worth, their responses proved more balanced and open.

REDUCING RESISTANCE TO DISCONFIRMING EVIDENCE

Not only may people interpret ambiguous information in a biased manner, they may also resist evidence that outright invalidates their preexisting beliefs. A second set of studies tested the hypothesis that responses to such evidence would be less defensive if self-integrity was secured by an affirmation of an alternative source of self-worth. If the motivation to maintain self-integrity is thus satisfied, people should be more willing to give up a cherished belief when reason or experience dictates they should.

In one study (Cohen et al., 2000), devout opponents and proponents of capital punishment were presented with a persuasive scientific report that contradicted their beliefs about the death penalty's effectiveness as a deterrent for crime. As in past research (Lord et al., 1979), participants under normal circumstances exhibited a disconfirmation bias. They found flaws in the methodology of the studies reported, they suspected bias on the part of the authors of the report, and they persisted in their attitudes toward capital punishment. By contrast, the responses of participants who received an affirmation of a valued self-identity (by writing about a personally important value, or by being provided with positive feedback on an important skill) proved more favorable. Self-affirmed participants were less critical of the reported re-

search, they suspected less bias on the part of the authors, and they even changed their overall attitudes toward capital punishment in the direction of the report they read. The affirmation freed people from the desire to protect self-worth, and thus enabled them to review the report less defensively.

When information concerns a risky health behavior, resistance can prove particularly costly. To encourage people to change their risky behavior, many health educators try to convince them that their actions pose serious health risks. However, these health campaigns can threaten individuals' views of themselves as healthy and adaptive. Once again, the ironic consequence may be resistance to information that, if acted upon, could ultimately improve their health and adaptiveness.

In one study (Sherman et al., 2000), sexually active undergraduates watched an educational video suggesting that their sexual behavior could put them at risk for HIV. Half received a self-affirmation prior to watching the video; the others did not. Although nonaffirmed participants tended to resist the presented information, affirmed participants responded by acknowledging their potential risk for contracting AIDS. The effects of affirmation also went beyond perceived personal risk and were seen in actual health behaviors. Whereas 25% of nonaffirmed individuals purchased condoms after viewing the video, 50% of affirmed participants did so. Defensiveness in response to health-risk messages may thus arise because such messages, by contradicting individuals' beliefs about their freedom from risk, inadvertently threaten self-worth. However, a small intervention can buffer people against this threat and thus promote desirable health behavior.

These findings challenge assumptions in social and health psy-

chology concerning the intractability of defensive biases. Past approaches (e.g., Weinstein & Klein, 1995) have attempted to reduce perceived invulnerability to health risk by using informational strategies—for example, by presenting facts about risk factors. However, such strategies are generally ineffective (Weinstein & Klein, 1995). One reason, it appears, is that informational approaches fail to address the ego-protective motivations underlying defensive biases in response to health-risk information. In contrast, the motivational approach featured in the research just summarized accomplishes precisely this, and in so doing, reduces bias and prompts positive change in health behavior.

REDUCING DEFENSIVE RESPONSES TO SUCCESS AND FAILURE

After the 2000 Summer Olympics, Bela Karolyi commented on the two gymnastics teams he had been associated with—the defeated United States team he advised in 2000 and the victorious Romanian team he had last coached in 1976. With respect to the defeated U.S. team, he said, "This new generation does not have quite the fight or the will or the work ethic of the previous generation" (Pucin, 2000, p. U5). But with respect to the winning Romanian team, he said, "I am very proud. Especially since I had a little part of their road to success" (Pucin, 2000, p. U5). The self-serving tendencies Karolyi displayed also occur at the group level—losing teams tend to blame their defeat on external factors, such as luck, whereas winning teams tend to attribute their victories to internal factors, such as their own effort, strategy, and skill (e.g., Lau & Russell, 1980). Although self-serving responses to success

and failure may help to maintain self-worth, they could, like the biases we discussed earlier, limit openness to potentially beneficial information. Losers risk ignoring areas of weakness, while winners risk resting on their laurels.

One series of studies (Sherman & Kim, 2002) investigated whether this bias would be attenuated by a self-affirmation. Indirectly securing self-worth should decrease the ego's need to interpret events in a self-serving manner. In one study, athletes who had just completed an intramural volleyball game assessed the extent to which each of a series of factors contributed to their team's victory or defeat. As in past research (Lau & Russell, 1980), winners made more internal attributions for their victories than losers did for their defeats. However, among athletes who had reflected on an important value irrelevant to athletics, this self-serving bias was attenuated. Once again, people were more open to important but threatening information when their self-worth was affirmed.

FINAL THOUGHTS

An important unresolved issue concerns the effect of culture on affirmation processes. It seems likely that the effects of affirmation in collectivist cultures will be different from the effects observed in individualist cultures. One possibility suggested by Heine and Lehman (1997) is that members of collectivist cultures may be less motivated to protect self-integrity because their culture places less emphasis on maintaining a positive self-image. However, it is also possible that members of collectivist cultures are just as motivated as members of individualist cultures to protect self-integrity, but that they would be more responsive to collectivist affirmations (e.g., of so-cial relationships) than to individualist affirmations (e.g., of personal values; see Heine & Lehman, 1997).

Another unresolved issue concerns the processes mediating the effect of self-affirmation. According to our analysis, affirmations make salient alternative sources of self-worth, allowing people to accept threatening information without feeling that their overall self-integrity is in jeopardy. One important question involves whether self-affirming thoughts must be consciously activated or whether affirmations can be nonconscious. Research using primes administered outside of conscious awareness suggests that self-affirming thoughts need not be conscious (Cohen, 2002). Compared with participants primed on failure words, participants primed on self-affirming, success words proved less likely to rationalize failure on a test by defensively devaluing the importance of the ability being assessed. Another question concerns the mediating role of mood. Affirmation may elevate mood, which may, in turn, increase openness to persuasion. In two separate studies (Cohen et al., 2000; Sherman et al., 2000), however, we found no effects of affirmation on self-reported mood. Moreover, a study by Correll, Spencer, and Zanna (2001) found that the effect of affirmation on increasing openness to persuasion was more pronounced if the arguments were strong rather than weak—a result inconsistent with typical mood effects observed in research on persuasion. Nevertheless, it is possible that nonconscious affective processes mediate the effects of affirmation (see Koole, Smeets, van Knippenberg, & Dijksterhuis, 1999). Social-cognitive methods, which permit systematic manipulation and measurement of nonconscious cognitive and affective variables, could help to disentangle relevant mediating processes further.

At a theoretical level, our research supports a motivational ac-count of defensive biases. Such biases arise, in part, from pressures to maintain self-integrity. Although they may be adaptive in some respects—helping to maintain self-esteem and well-being (Taylor, 1989)—they may also raise barriers to change and to the resolution of conflict. Affirmations of alternative sources of self-worth, however, can sharply attenuate defensive biases, and encourage attitude and behavior change in potentially threatening or contentious domains. Such self-affirmations, it seems, allow people to evaluate evidence on the basis of its merits rather than its correspondence with their beliefs, desires, and vested interests.

Recommended Reading

Aronson, J., Cohen, G.L., & Nail, P.R. (1999). Self-affirmation theory: An update and appraisal. In E. Harmon-Jones & J. Mills (Eds.), *Cognitive dissonance: Progress on a pivotal theory in social psychology* (pp. 127–147). Washington, DC: APA Books.

Crocker, J., & Wolfe, C.T. (2001). Contingencies of self-worth. *Psychological Review, 108*, 593–623.

Dunning, D. (2001). On the motives underlying social cognition. In N. Schwarz & A. Tesser (Eds.), *Blackwell handbook of social psychology: Vol. 1. Intraindividual processes* (pp. 348–374). New York: Blackwell.

Notes

1. Address correspondence to David Sherman, Department of Psychology, UCLA, 1282A Franz Hall, 405 Hilgard Ave., Los Angeles, CA 90095-1563, e-mail: sherman@psych.ucla.edu, or to Geoffrey Cohen, Department of Psychology, Yale University, 2 Hillhouse Ave., Box 208205, New Haven, CT 06520-8205, e-mail: geoffrey.cohen@yale.edu.

2. Coffee drinkers need not worry; the coffee-drinking/breast cancer

studies presented to participants in the research we summarize were fabricated.

References

Cohen, G.L. (2002). [Priming effects on defensive rationalization]. Unpublished raw data.

Cohen, G.L., Aronson, J., & Steele, C.M. (2000). When beliefs yield to evidence: Reducing biased evaluation by affirming the self. *Personality and Social Psychology Bulletin, 26*, 1151–1164.

Correll, J., Spencer, S., & Zanna, M. (2001). *Motivated resistance to counterattitudinal arguments: The effects of affirmation, argument strength, and attitude importance*. Unpublished manuscript, University of Waterloo, Waterloo, Ontario, Canada.

Heine, S.J., & Lehman, D.R. (1997). Culture, dissonance, and self-affirmation. *Personality and Social Psychology Bulletin, 23*, 389–400.

Koole, S.L., Smeets, K., van Knippenberg, A., & Dijksterhuis, A. (1999). The cessation of rumination through self-affirmation. *Journal of Personality and Social Psychology, 77*, 111–125.

Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology, 53*, 636–647.

Lau, R.R., & Russell, D.R. (1980). Attributions in the sports pages. *Journal of Personality and Social Psychology, 39*, 29–38.

Lord, C.G., Ross, L., & Lepper, M.R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*, 2098–2109.

Nisbett, R.E., & Ross, L.D. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.

Pucin, D. (2000, September 20). Pixies fall from grace. *Los Angeles Times*, pp. U1, U5.

Sherman, D.A.K., Nelson, L.D., & Steele, C.M. (2000). Do messages about health risks threaten the self: Increasing the acceptance of threatening health messages via self-affirmation. *Personality and Social Psychology Bulletin, 26*, 1046–1058.

Sherman, D.K., & Kim, H.S. (2002). *Increasing attributions of personal responsibility in sports competitions via self-affirmation*. Unpublished manuscript, University of California, Los Angeles.

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.

Taylor, S.E. (1989). *Positive illusions: Creative self-deception and the healthy mind*. New York: Basic Books.

Weinstein, N.D., & Klein, W.M. (1995). Resistance of personal risk perceptions to debiasing interventions. *Health Psychology, 14*, 132–140.

What We Know Now About Bias and Intergroup Conflict, the Problem of the Century

Susan T. Fiske[1]

Department of Psychology, Princeton University, Princeton, New Jersey

Abstract

After nearly a century's study, what do psychologists now know about intergroup bias and conflict? Most people reveal unconscious, subtle biases, which are relatively automatic, cool, indirect, ambiguous, and ambivalent. Subtle biases underlie ordinary discrimination: comfort with one's own in-group, plus exclusion and avoidance of out-groups. Such biases result from internal conflict between cultural ideals and cultural biases. A small minority of people, extremists, do harbor blatant biases that are more conscious, hot, direct, and unambiguous. Blatant biases underlie aggression, including hate crimes. Such biases result from perceived intergroup conflict over economics and values, in a world perceived to be hierarchical and dangerous. Reduction of both subtle and blatant bias results from education, economic opportunity, and constructive intergroup contact.

Keywords

bias; stereotyping; prejudice; discrimination; intergroup conflict

People typically seek other people who are similar to themselves, being comfortable with others they perceive as members of their own in-group. From comfort follows, at best, neglect of people from out-groups and, at worst, murderous hostility toward out-groups perceived as threatening the in-group. Biases do vary by degree, and the psychologies of moderate and extreme biases differ considerably. Well-intentioned moderates reveal bias more subtle than the rants and rampages of extremists. By some counts, 80% of Western democratic populations intend benign intergroup relations but display subtle biases. In contrast, blatantly biased extremists are completely out-front. Although estimated to be a minority (perhaps 10%), they are salient, vocal, and dangerous.

After nearly a century's study, social psychology knows a lot about both forms of bias. Stereotyping, prejudice, and discrimination reflect, respectively, people's cognitive, affective, and behavioral reactions to people from other groups (Fiske, 1998). All constitute *bias*, reacting to a person on the basis of perceived membership in a single human category, ignoring other category memberships and other personal attributes. Bias is thus a narrow, potentially erroneous reaction, compared with individuated impressions formed from personal details.

SUBTLE BIAS AMONG WELL-INTENTIONED MODERATES

Automatic, Unconscious, and Unintentional

The big news from two recent decades of research: Bias is most often