Sources of Implicit Attitudes

Laurie A. Rudman

Rutgers University

ABSTRACT—*Response latency measures have yielded an explosion of interest in implicit attitudes. Less forthcoming have been theoretical explanations for why they often differ from explicit (self-reported) attitudes. Theorized differences in the sources of implicit and explicit attitudes are discussed, and evidence consistent with each theory is presented. The hypothesized causal influences on attitudes include early (even preverbal) experiences, affective experiences, cultural biases, and cognitive consistency principles. Each may influence implicit attitudes more than explicit attitudes, underscoring their conceptual distinction.*

KEYWORDS—*implicit attitudes; implicit social cognition; response latency measures*

Attitude researchers have long been wary of taking people's reports of their own attitudes at face value, particularly when the topics being considered impinge on people's morality. Prejudice and attitudes toward immoral (e.g., cheating) or illegal (e.g., substance abuse) behaviors are but a few examples of such topics. Response latency measures, which yield evaluations that are unlikely to be controlled, have been heralded because they override the obvious problem of distortion. This is because they are taken from reaction time tasks that measure people's attitudes or beliefs indirectly (i.e., without asking people how they feel or think). That is, people's attention is focused not on the attitude object, but on performing an objective task, and attitudes are then inferred from systematic variations in task performance. For example, in the Implicit Association Test, automatic pro-White bias is indicated when people show faster performance categorizing pleasant words and Whites (and unpleasant words and Blacks) together, compared with categorizing unpleasant words and Whites (and pleasant words and Blacks) together. Thus, implicit attitudes can be characterized as the automatic association people have between an object and evaluation (whether it is good or bad). By contrast, explicit attitudes may reflect more thoughtful or deliberative responding.

More substantively, response latency measures have also led to increased interest in potential theoretical differences between implicit and explicit attitudes. Although there have been no formal frameworks (cf. Fazio & Olson, 2003), there have been a few attempts to conceptually distinguish what is being measured when people report their attitudes from what is measured by response latency. In this review, I discuss four factors that have been theorized to influence implicit more than explicit attitudes. Although preliminary, the evidence suggests that automatic and controlled evaluations stem from different sources and, therefore, should be conceptualized as distinct constructs.

EARLY EXPERIENCES

A prominent conception is that implicit attitudes stem from past (and largely forgotten) experiences, whereas explicit attitudes reflect more recent or accessible events (Greenwald & Banaji, 1995). In a study supporting this hypothesis (Rudman & Heppen, 2001), smokers' implicit attitudes toward their habit covaried with their earliest experiences with smoking, which were mainly unpleasant (e.g., aversion to tobacco smoke and nausea from their first cigarettes). Thus, automatic attitudes were negative if early experiences were unpleasant. By contrast, smokers' explicit attitudes covaried with their recent experiences toward smoking, which were mainly positive (e.g., drinking coffee and smoking with friends). Thus, self-reported attitudes were negative if recent experiences were unpleasant. Differences in the underlying sources of smokers' implicit and explicit attitudes helps to explain why they were only weakly related.

A logical extension of Greenwald and Banaji's (1995) argument is that developmental events may inform implicit more than explicit attitudes. Much of what is learned early in life is preverbal and taught indirectly. These lessons form the foundation on which later learning is built and may also serve as a nonconscious source for related evaluations and actions. Goodwin and I obtained results consistent with this possibility when we investigated whether early (even preverbal) attachment to maternal caregivers was associated with people's gender-related attitudes (Rudman & Goodwin, 2003). First, people raised primarily by their mothers implicitly preferred women to men. Second, people implicitly favored women if they automatically preferred their mothers to their fathers. By contrast, explicit attitudes toward parents and gender were not related.

Finally, in three experiments (Rudman & Heppen, 2003), women who possessed an automatic association between romantic partners and chivalric roles (e.g., White Knight, Prince Charming) reported less interest in personal power, including economic and educational achievement, than woman who did not have this automatic association. By contrast, explicit romantic fantasies did not covary with explicit power-related variables. Because women are socialized early and often to view men as their heroes and rescuers (e.g., through romantic fairy tales), these findings indirectly support the hypothesis that developmental events can inform automatic mental habits.

Address correspondence to Laurie A. Rudman, Department of Psychology, Tillett Hall, Rutgers University, 53 Avenue E, Piscataway, NJ 08854-8040; e-mail: rudman@rci.rutgers.edu.

AFFECTIVE EXPERIENCES

It is also possible that implicit attitudes are more sensitive to affective experiences than are explicit attitudes. For example, Phelps et al. (2000) found that estimates of implicit (but not explicit) prejudice positively covaried with activation in a brain structure called the amygdala in Whites exposed to photos of Blacks. Because the amygdala is implicated in the control of affective responses, these results suggest that implicit attitudes may stem from automatic emotional reactions to stimuli, whereas explicit attitudes may be "cooler" (more cognitively controlled).

In addition, my colleagues and I found that Whites who volunteered for diversity education showed reduced anti-Black attitudes, both implicit and explicit, at the end of the course. However, changes in the two kinds of attitudes were only weakly associated. Further examination revealed that reductions in implicit attitudes were linked to emotion-based predictors, including reduced fear of Blacks, increased friendships with Blacks, and liking for the African American professor who taught the course. By contrast, reductions in explicit attitudes covaried with students' increased awareness of bias and their desire to overcome their own prejudice (i.e., "trying hard" to change). These findings suggest that changes in implicit attitudes may depend on emotional reconditioning, whereas changes in explicit attitudes may depend on more cognitive and motivational factors (Rudman, Ashmore, & Gary, 2001).

Finally, the sensitivity of implicit attitudes to priming effects (i.e., to the influence of contextual factors) has now been well established (Blair, 2002). Because priming manipulations are recent events, at first blush such findings appear to conflict with the hypothesis that early experiences impact implicit more than explicit attitudes. However, the two views can be reconciled if affect comes into play during the priming manipulation. Learning about admired Blacks and criminal Whites, mentally imagining heroic women, and listening to rap music have all been shown to modify implicit associations. It is possible that these effects were due, at least in part, to the feelings aroused by the stimuli. Likewise, it is conceivable that the studies that found past events influenced implicit attitudes more than explicit attitudes obtained these results precisely because the events were emotional (e.g., aversive experiences with smoking, maternal bonding, and romantic fantasies). Although speculative, the possibility that affect accounts for the influence of both recent and past experiences on implicit attitudes seems worthy of pursuit.

CULTURAL BIASES

The third possibility is that implicit attitudes are more influenced by one's cultural milieu than explicit attitudes are (Devine, 1989). For example, Greenwald, McGhee, and Schwartz (1998, Experiment 2) found that Korean and Japanese American students showed greater automatic in-group bias to the extent they were immersed in their ancestors' culture (e.g., spoke the language). The linkage between attitudes and culture was less evident using a self-report measure of attitudes.

Further, it has been shown repeatedly that Blacks and Whites alike possess more anti-Black bias on implicit measures, compared with self-reports (e.g., Nosek, Banaji, & Greenwald, 2002). Although the pattern for Blacks is provocative, it is consistent with system-

justification theory's argument that minorities nonconsciously rationalize their lower status by internalizing society's negative view of their group (Jost & Banaji, 1994). My colleagues and I (Rudman, Feinberg, & Fairchild, 2002) tested this hypothesis using minority groups whose relative status (based on explicit ratings from an independent sample) ranged from low (poor, overweight) to high (Asians, Jews). The results supported system-justification theory. First, the lower their cultural status, the more minorities implicitly favored the dominant out-group. In fact, poor and overweight participants showed significant preference for rich and slim out-group members, respectively. Second, participants were asked to report their group's relative status, and these ratings also covaried with their implicit attitudes. For example, Jews who ranked Christians as higher in status than Jews tended to automatically associate Christians with positive attributes and Jews with negative attributes. In contrast to these results, minorities showed robust explicit in-group bias, which was unrelated to their status.

More dramatically, Livingston (2002) found that social standing had opposite influences on Blacks' automatic and self-reported in-group bias. Specifically, Blacks who perceived that Whites disliked their group showed stronger automatic pro-White bias, but at the same time, stronger pro-Black bias in their self-reports, compared with Blacks who perceived that Whites liked their group. When the same variable pulls implicit attitudes in one direction and explicit attitudes in another, their conceptual distinction is strongly supported.

Finally, high-status groups (e.g., Whites, Christians, slim people, rich people) routinely show stronger implicit in-group bias than do low-status groups, but again, this is a function of their relative status (whereas explicit in-group bias is not; Rudman et al., 2002). Thus, for members of dominant and minority groups alike, societal evaluations appear to have an assimilative effect on automatic (but not controlled) attitudes, suggesting that cultural biases inform implicit attitudes more than explicit attitudes. Because learning about one's place in the world is likely to occur early (and often) in life, and is likely to be emotionally charged, the influence of cultural biases on implicit attitudes may be reconcilable with the influence of early and affective experiences.

COGNITIVE CONSISTENCY PRINCIPLES

A venerable principle in social psychology is that people prefer consonant (as opposed to dissonant) evaluations of related attitude objects. For example, according to this principle, if I like myself and I am female, then I should also like women. This prediction means that cognitive consistency should be observed among the variables of self-esteem, gender identity, and gender attitude. In a compelling demonstration suggesting that automatic and controlled evaluations stem from different causes, implicit attitudes, identity, self-esteem, stereotypes, and self-concept conformed to cognitive consistency principles, whereas self-reports of these same constructs did not (an observation that led the development of the unified theory of implicit social cognition; Greenwald et al., 2002). The general pattern of results the unified theory predicts for implicit measures can be characterized as "If I am Y and I am X, then X is also Y," where Y represents evaluation and X represents group membership. In five experiments, this pattern of results was obtained using the Implicit Association Test. For example, Whites who showed high self-esteem and who identified with their ethnicity also preferred Whites to

Blacks. Thus, the logic underlying implicit attitudes was "If I am good and I am X, then X is also good." A similar pattern emerged when implicit stereotypes and self-concept were measured, rather than attitudes and self-esteem. For example, men and women who associated themselves with warmth (or power) also associated warmth (or power) with their own gender, provided they identified with their gender; self-report measures did not conform to this pattern. Identical findings were found using an academic (math-arts) gender stereotype, but again, only with implicit estimates of self-concept, stereotypes, and gender identity. By uncovering cognitive balance at the automatic level, research that supports the unified theory underscores important differences in the sources of implicit and explicit constructs.

The unified theory may be reconciled with other theories that distinguish sources of implicit and explicit attitudes. First, the theory converges with the hypothesis that cultural milieu biases implicit attitudes. Societal evaluations clearly influence implicit in-group appraisal, which contributes to self-appraisal when in-group identification is strong. Second, affect may inform the unified model by means of evaluative links that involve the self. Given that people likely do not view themselves impartially, emotional self-appraisals may spill over into automatic (more than self-reported) evaluations. For example, early and affective lessons learned about the self may shape one's implicit appraisals of other objects that are (or are not) connected to the self. Interestingly, the resulting implicit structure can be counterstereotypical (e.g., "If I am warm and I am male, then men are warmer than women"). In this way, automatic self-appraisals may counter the influence of culture on implicit associations.

FUTURE DIRECTIONS

In sum, the preliminary evidence indicates that early and affective experiences may influence automatic evaluations more than explicit attitudes. In addition, there is growing evidence that systemic, culturally held appraisals can bias people's automatic evaluations irrespective of their personal opinion. Finally, only implicit (not explicit) evaluations appear to be sensitive to cognitive consistency principles. By better understanding disparities in the underlying causes of implicit and explicit attitudes, psychologists can begin to formulate more sophisticated frameworks for conceptualizing them. At the very least, the observation that they stem from different sources suggests they should be viewed as theoretically distinct.

The argument that implicit and explicit attitudes are conceptually distinct can help to explain why the two types of attitudes are often dissociated, and why response latency measures (sometimes) predict behaviors better than self-reports. Although the untrustworthiness of self-reports is often blamed for these findings, the picture is more complex than that. For one thing, implicit and explicit attitudes sometimes correlate well. This is particularly true when attitude objects are noncontroversial (e.g., politicians, academic subjects, and vegetarianism), but even measures of implicit and explicit prejudice sometimes converge. Moreover, implicit and explicit attitudes can be dissociated even for noncontroversial objects (e.g., flowers, insects, apples, candy bars). Thus, a challenge for future research is to uncover the variables that determine when implicit and explicit attitudes converge. To meet this challenge, researchers should go beyond the controllability of self-reports as an explanation for the weak convergence between implicit and explicit attitudes because this ex-

planation assumes that the underlying evaluation is the same for both types of attitudes (and people are either unable or unwilling to report their "true" attitude). If automatic and controlled evaluations stem from different sources, their underlying valence may dramatically differ. That is, implicit attitudes may be unfavorable despite favorable explicit attitudes, and vice versa. Attending to source differences provides a rationale for deeming both kinds of evaluations as genuine, albeit limited in their ability to encompass the range of human responses to attitude objects.

Future work should also focus on the conditions under which implicit attitudes predict behavior better than explicit attitudes. It has been proposed that automatic and controlled evaluations best predict spontaneous and deliberative actions, respectively. Although there is some evidence that implicit attitudes are linked to involuntary behaviors (e.g., eye blinking), the larger picture is more complicated. First, implicit attitudes often influence deliberative actions, including choosing which consumer products to purchase, volunteering for leadership roles, using condoms, and discriminating against job applicants who are minority-group members. Second, there is substantial evidence, some of it presented here, that controlled responses covary with implicit attitudes (e.g., reports of early experiences covaried with implicit attitudes toward smoking, explicit predictions of status predicted implicit in-group bias, and self-reported interest in power correlated with implicit romantic fantasies). And finally, as already noted, implicit and explicit attitudes sometimes converge. Thus, although implicit attitudes might be the best predictor of spontaneous actions, they are also capable of predicting a large array of controlled behaviors. As opposed to taking a purely "process-matching" approach to predicting behavior, researchers should consider additional factors, including the extent to which the situation increases the salience of implicit attitudes and their relevance to the behavior or judgment at hand (see also Fazio & Olson, 2003).

Focusing on the automatic versus controlled nature of implicit and explicit attitudes has been fruitful, but may mask other ways in which they differ, including their underlying causes. I have presented four factors known to influence implicit attitudes more than explicit attitudes, and suggested how they might overlap. Attending to differences in sources should promote integrative theoretical frameworks that differentiate the two kinds of attitudes. It should also aid in identifying factors that modify each, variables that determine when implicit and explicit attitudes converge, and conditions that promote their utility in predicting behavior.

Recommended Reading

Banaji, M.R. (2001). Implicit attitudes can be measured. In H.L. Roediger & J.S. Nairne (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 117–150). Washington, DC: American Psychological Association.

Greenwald, A.G., Banaji, M.R., Rudman, L.A., Farnham, S.D., Nosek, B.A., & Mellott, D.S. (2002). (See References)

Wilson, T.D., Lindsey, S., & Schooler, T.Y. (2000). A model of dual attitudes. *Psychological Review, 107,* 101–126.

REFERENCES

Blair, I.V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*, 242–261.

Devine, P.G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5–18.

Fazio, R.H., & Olson, M.A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297–327.

Greenwald, A.G., & Banaji, M.R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*, 4–27.

Greenwald, A.G., Banaji, M.R., Rudman, L.A., Farnham, S.D., Nosek, B.A., & Mellott, D.S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109*, 3–25.

Greenwald, A.G., McGhee, D.E., & Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464–1480.

Jost, J.T., & Banaji, M.R. (1994). The role of stereotyping in system-justification and the production of false-consciousness. *British Journal of Social Psychology, 33*, 1–27.

Livingston, R.W. (2002). The role of perceived negativity in the moderation of African Americans' implicit and explicit racial attitudes. *Journal of Experimental Social Psychology, 38*, 405–413.

Nosek, B.A., Banaji, M.R., & Greenwald, A.G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 6*, 101–115.

Phelps, E.A., O'Connor, K.J., Cunningham, W.A., Gatenby, J.C., Funayama, E.S., Gore, J.C., & Banaji, M.R. (2000). Amygdala activation predicts performance on indirect measures of racial bias. *Journal of Cognitive Neuroscience, 12*, 729–738.

Rudman, L.A., Ashmore, R.D., & Gary, M.L. (2001). "Unlearning" automatic biases: The malleability of implicit stereotypes and prejudice. *Journal of Personality and Social Psychology, 81*, 856–868.

Rudman, L.A., Feinberg, J.M., & Fairchild, K. (2002). Minority members' implicit attitudes: Ingroup bias as a function of group status. *Social Cognition, 20*, 294–320.

Rudman, L.A., & Goodwin, S.A. (2003). *Gender differences in automatic in-group bias: Why do women like women more than men like men?* Manuscript submitted for publication.

Rudman, L.A., & Heppen, J. (2001). *The smoking gun: Implicit and explicit attitudes toward smoking*. Manuscript submitted for publication.

Rudman, L.A., & Heppen, J. (2003). Implicit romantic fantasies and women's interest in personal power: A glass slipper effect? *Personality and Social Psychology Bulletin, 29*, 1357–1370.