

## Implicit Prejudice and Stereotyping: How Automatic Are They? Introduction to the Special Section

Patricia G. Devine  
University of Wisconsin—Madison

This special issue of the *Journal of Personality and Social Psychology: Attitudes and Social Cognition* addresses issues of the measurement and the malleability of implicit prejudice and stereotypes. The findings raise fundamental questions about the assumptions underlying the assessment of implicit prejudice, particularly with regard to the widely used Implicit Association Test (A. Greenwald, D. McGhee, & J. Schwartz, 1998) and the assumption of extant models of prejudice and stereotyping that implicit biases are automatically and invariantly activated when perceivers come in contact with members of stigmatized groups. Several of the articles show that contextual manipulations produce reductions in implicit manifestations of prejudice and stereotyping. The articles in this issue, in challenging conventional wisdom, are thought provoking and should be generative in the field's ongoing efforts to understand the role of implicit (and explicit) processes involved in prejudice and stereotyping.

In recent years, there has been a veritable explosion of work on the nature and assessment of implicit components of prejudice and stereotyping. Over the last decade or so, a great many studies have revealed that prejudice and stereotypes can operate without the conscious intent or awareness of social perceivers. Even those who consciously renounce prejudice have been shown to have implicit or automatic biases that conflict with their nonprejudiced values that may disadvantage the targets of these biases. Theorists have become consumed with questions concerning the pervasiveness of such biases and when and how such biases manifest themselves. Indeed, few issues have so completely captured the interest and imagination of many well-established researchers as well as those who are new to the discipline. The reasons for the widespread interest range from the obvious to the more subtle. Members of stigmatized groups have been disadvantaged owing to prejudice and stereotyping on the part of the nonstigmatized. Thus, from a practical perspective, understanding the nature of implicit prejudice and stereotyping could lead to effective interventions to reduce or eliminate their pernicious effects. Theoretically, understanding the nature of implicit prejudice and stereotyping is just as exciting. That is, studying such socially significant issues in theoretically sophisticated and methodologically rigorous ways may unlock some of the puzzling aspects of how information is represented in memory, accessed, and used in social judgment and behavior.

A perhaps less obvious reason for the high level of interest in the nature of implicit processes is that it has effectively drawn out the

creativity and imaginativeness of social psychologists. Nowhere is this creativity and imagination more evident than in the development of new tools to study the existence and operation of such processes. Whereas such unconscious or implicit processes were once the domain of psychoanalysis and open only to speculation, the development of effective tools to uncover and study implicit, automatic, unconscious processes makes such processes tractable and open to careful scientific scrutiny. Greenwald, McGhee, and Schwartz (1998) recently introduced one such tool, the Implicit Association Test (IAT), that purports to measure implicit biases.

The measure has been widely applied in the context of exploring the nature of implicit biases against a variety of out-groups (e.g., race, gender, and age). Few measurement tools have received greater empirical attention than the IAT. Part of the reason for the widespread use of the IAT is that Greenwald and his colleagues (Greenwald et al., 1998) have generously made the measure readily available to others; in addition, it is easy to use and reliably produces strong effects. These features make it appealing as a measurement tool. Another reason for the high level of interest in the IAT is that social psychologists have long been in search of measures that could bypass or otherwise circumvent social desirability biases associated with the measurement of prejudice. Greenwald et al. (1998) have suggested that the IAT is one such measure.

Of course, any new measure is only useful to the extent that it validly measures what it purports to measure. Although many have unquestioningly accepted the IAT as a measure of implicit prejudice or stereotyping, others take a different view. Questions concerning the validity of new measures are inevitable and are healthy for the discipline. Although the IAT offers great potential for addressing the nature and functioning of implicit biases, the widespread use of it as a tool for measuring prejudice and stereotyping

---

Correspondence concerning this article should be addressed to Patricia G. Devine, Department of Psychology, University of Wisconsin—Madison, 1202 West Johnson Street, Madison, Wisconsin 53706. Electronic mail may be sent to [pgdevine@facstaff.wisc.edu](mailto:pgdevine@facstaff.wisc.edu).

quite naturally leads to questions concerning the validity and meaning of responses on the measure. Questions of this sort will likely lead to a better understanding of the IAT as a measurement tool and of the meaning of implicit biases more generally. At the same time, many researchers have been using the IAT as a strategy to assess implicit prejudices and stereotypes and to show that extant theories of stereotyping and prejudice that assume such biases are automatically and invariantly activated in the presence of stigmatized group members may be in need of revision. Therefore, theoretical assumptions central to our theories of stereotyping and prejudice are being challenged, in large part on the basis of findings using the IAT.

This special issue of the Attitudes and Social Cognition section of the *Journal of Personality and Social Psychology* (JPSP) did not arise in the ordinary fashion of special issues. There was no preplanned theme for a special issue, and there was no call for papers. What there was, however, was a great deal of interest and research activity on the nature of biases revealed by the IAT (among other measures) and the possibility that such biases are not as immutable as suggested by extant theories. In a very short period of time, several articles were submitted to JPSP addressing these basic questions. Some of the articles raised fundamental questions concerning exactly what is measured by the IAT and offered interpretations that differed from the preferred interpretation of Greenwald et al. (1998) that the IAT is capable of revealing implicit prejudice. Other articles addressed the question of the extent to which implicit biases are immutable. These articles each attempted to show that situational or contextual manipulations produce diminutions of implicit or automatic race or gender biases. If such spontaneous processes are so easily moderated, there are likely to be serious implications for models that have argued that automatic processes are well learned and are, as a result, difficult to overcome. Taken together, these articles raise fundamental questions concerning the meaning of and the immutability of implicit prejudice and stereotyping.

Each article was reviewed independently, and in each case the reviewers' reactions overlapped quite substantially. The reviewers found the questions to be interesting and important. They found the findings to be thought provoking and potentially exciting. As well, in each case, they were not convinced that the theoretical processes underlying the effects reported were sufficiently well documented. This situation created an interesting dilemma for the editors handling these articles.<sup>1</sup> Most JPSP articles require a clear articulation of the processes underlying the phenomena of interest. Indeed, this is typically the central objective of articles published in JPSP. Nevertheless, the findings reported in this set of articles are important if only because they challenge in some compelling ways conventional wisdom concerning the nature of implicit or automatic prejudice and stereotyping. They are important as well because, although they may raise more questions than they answer, the questions raised are themselves important and will, in my estimation, very likely be generative. Indeed, it was striking that so many laboratories quite independently produced a consistent pattern of findings concerning the malleability of implicit biases. The fact that these various empirical efforts were conducted simultaneously reflects the fact that the issue is central to current developments in the study of stereotyping and prejudice.

Because of the provocative nature of these findings and their potential to inform the field's understanding of the role of implicit

processes in stereotyping and prejudice, I decided to pull these articles together for a special issue of JPSP: Attitudes and Social Cognition. I believe the fundamental finding, replicated in different laboratories and with different manipulations of context, is important and will stimulate additional research that will be important for the development of theory about prejudice and stereotyping and their multifaceted components. Despite the high level of activity on the IAT and other implicit bias measures, we currently know very little about these measures and how they will ultimately contribute to our understanding of prejudice. I believe that research adding to this knowledge base will facilitate the development of the field's understanding of such measures, their relation to explicit measures, and, ultimately, their relation to behavior. Therefore, this special issue makes a valuable contribution to the prejudice and stereotyping literatures specifically and to the literature addressing implicit and (perhaps) explicit social cognition more generally.

The articles by Brendl, Markman, and Messner (2001) and by Karpinski and Hilton (2001) raise fundamental questions about the IAT and what it measures. Each of these authors seriously questions the extent to which the IAT measures prejudice. Brendl et al. suggest that response criterion shifts can affect the IAT and, therefore, its utility as a measure of implicit bias. Karpinski and Hilton argue that the IAT measures well-learned environmental associations rather than endorsed negative reactions concerning attitude objects. In one way or another, these authors suggest that, contrary to prevailing assumptions, the IAT does not measure well-learned, established biases against stigmatized groups members. Ashburn-Nardo, Voils, and Monteith (2001) add fuel to such questions when they show in a compelling fashion that the IAT appears to measure in-group preferences even in minimal group settings, when familiarity with in-groups and out-groups is low. They argue that the automatic nature of evaluative differentiation between in-groups and out-groups suggests readiness of the human mind to receive and accept biases favoring the in-group. An important finding from this set of studies is that clear evidence of bias is present even when participants have no preexisting experience with or bias toward the in-groups or out-groups. Again, questions concerning the exact meaning of IAT responses arise.

The other articles in this issue each show the effect of a variety of situational or contextual manipulations in moderating the magnitude of implicit biases. It is worth noting that each of the authors begins with the assumption that the IAT (or other implicit measures) assesses prejudice (or stereotyping). This assumption stands in stark contrast to the conclusions of Brendl et al. (2001) and Karpinski and Hilton (2001). Whether these measures assess prejudice or not, it is clear that context manipulations produce replicable patterns of moderation of implicit biases. Dasgupta and Greenwald (2001) show that thinking about admired out-group members (e.g., Tiger Woods) and disliked in-group members (i.e., Jeffrey Dahmer) leads to a diminution of implicit race bias on the

<sup>1</sup> The articles included in this special issue were processed by three different action editors: Arie W. Kruglanski, Patricia G. Devine, and Timothy D. Wilson. The Brendl et al. (2001) article was accepted by Arie W. Kruglanski at the end of his editorial term. Brendl et al. agreed to have the publication of the article delayed by a few months to allow inclusion in this special issue.

IAT, an effect that lasts up to 24 hr. Dasgupta and Greenwald reason that their manipulation may make salient positive exemplars of the out-group and affect the extent to which participants associate Blacks with evaluatively positive information. Wittenbrink, Judd, and Park (2001) explore context effects on the IAT by manipulating whether stigmatized others (e.g., Blacks) were encountered in positive (e.g., family barbecue) or negative (e.g., gang incident) settings, presumably making positive and negative components of the stereotype temporarily accessible. In a second study, they show the moderation of implicit biases within participants and on a different measure of implicit bias (i.e., sequential priming measure). These authors contend that such findings should not be surprising (even though they are at odds with conventional wisdom) because our cognitive representations of out-groups are likely to be complex and include both positive and negative information. What is revealed on implicit measures is determined by what knowledge is activated by situational conditions.

Blair, Ma, and Lenton (2001) explore the effects of mental imagery, a controllable mental process, in producing moderation of implicit gender stereotyping. In a series of five studies, Blair et al. show that imagining counterstereotypic women (e.g., strong women) reliably produces reduction in the magnitude of implicit stereotyping revealed on the IAT and two other measures of implicit gender stereotyping. This set of studies is provocative in that although they each show reductions of implicit gender stereotyping, the process by which this occurs appears to vary across the measures (i.e., increasing the accessibility of counterstereotypic information vs. decreasing the accessibility of stereotypic information). Blair et al. suggest that replication across measures, each of which involves different response requirements (e.g., reaction time vs. signal detection sensitivity), provides the opportunity to refine our understanding of the processes by which manipulations have their effect.

Lowery, Hardin, and Sinclair (2001) argue that situational variability in implicit biases results from ordinary social interaction and communication processes. They show, for example, that White participants show stronger negative implicit race biases (on a paper-and-pencil version of the IAT and on a priming measure) when the experimenter is White than when the experimenter is Black. Their preferred interpretation for this moderation of implicit race biases derives from shared reality theory, in which social tuning of attitudes occurs automatically to meet the social relationship demands of any given interaction (i.e., negative race bias would be less evident when one needs to regulate a social relationship with a Black individual). Although speculative, this approach places the (non)activation of race biases squarely in the context of social interactions in which, presumably, race biases could do the most damage. Moreover, the findings, as do many of the others discussed, show that implicit biases are more malleable than previous theorizing suggests. Rudman, Ashmore, and Gary

(2001) explore contact of a different sort, diversity training, for reducing implicit prejudices and stereotyping. Students enrolled in a prejudice and conflict seminar taught by a Black professor showed a reduction in implicit biases after participation in the course, as compared with control students (enrolled in prejudice irrelevant classes taught by a White professor or the Black professor who taught the prejudice course).

Key questions arise as to whether any of these effects reflect true reductions in implicit prejudice or stereotyping. These questions are especially important given the ongoing disagreements concerning the meaning of implicit measures. All of the explanations advanced appear to have a subtyping component to them, and whether activation of different aspects of underlying cognitive representations reflects change is certainly open to debate. Indeed, I believe the theoretical explanation for the types of findings found in these articles is "up for grabs" at this point. In many ways, however, I do not see this as a problem but rather as an opportunity. I personally have enjoyed reading these articles and thinking about the issues as well as the challenges they pose for the field. My hope is that the articles in this issue will stimulate additional work and lead to new insights concerning both the measurement and the meaning of implicit prejudice and stereotyping.

## References

- Ashburn-Nardo, L., Voils, C. I., & Monteith, M. J. (2001). Implicit associations as the seeds of intergroup bias: How easily do they take root? *Journal of Personality and Social Psychology, 81*, 789–799.
- Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology, 81*, 828–841.
- Brendl, C. M., Markman, A. B., & Messner, C. (2001). How do indirect measures of evaluation work? Evaluating the inference of prejudice in the Implicit Association Test. *Journal of Personality and Social Psychology, 81*, 760–773.
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 81*, 800–814.
- Greenwald, A., McGhee, D., & Schwartz, J. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464–1480.
- Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81*, 774–788.
- Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology, 81*, 842–855.
- Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology, 81*, 856–868.
- Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology, 81*, 815–827.