Journal of Applied Psychology
2009, Vol. 94, No. 3, 598–603

0021-9010/09/$12.00 DOI: 10.1037/a0014666

Transparency Should Trump Trust: Rejoinder to McConnell and Leibold (2009) and Ziegert and Hanges (2009)

Hart Blanton
Texas A&M University

James Jaccard
Florida International University

Jonathan Klick
University of Pennsylvania

Barbara Mellers
University of California, Berkeley

Gregory Mitchell
University of Virginia

Philip E. Tetlock
University of California, Berkeley

The rebuttals offered by the authors whose data we reanalyzed (A. R. McConnell & J. M. Leibold, 2001; J. C. Ziegert & P. J. Hanges, 2005) address secondary issues that do not alter our primary message: The evidence for the predictive validity of the race Implicit Association Test is too fragile to support the strong claims that have been made about the pervasiveness of prejudice and the linkages between Implicit Association Test scores and discriminatory behavior. Greater caution in both the legal and scientific communities is warranted. Most importantly, scientific research on implicit bias needs greater transparency and willingness to open raw data to critical scrutiny, not greater trust and deference among researchers.

Keywords: Implicit Association Test, implicit bias, predictive validity, discrimination, replication

None of the responses offered by McConnell and Leibold (2009) or Ziegert and Hanges (2009) undercuts the results of our reanalysis project or our message that there needs to be greater scrutiny of, and greater transparency in, research examining the predictive validity of the Implicit Association Test (IAT). Space constraints allow only brief clarifications and rejoinders.

Bad Research Applications, Not Bad Research

First, our central thesis was not that any particular study was done incorrectly; rather, it was that the larger body of research from which we drew the two studies does not have the empirical properties necessary for supporting the sweeping policy implications that some commentators have proposed (e.g., Kang & Banaji, 2006). Indeed, even if the two studies had been analytically flawless and the IAT was a reasonable predictor of the discrimination indicators used in those studies, it would be inappropriate for social scientists, legal scholars, and journalists to draw strong descriptive and prescriptive conclusions from evidence with such obvious limitations, including the failure to consider the replicability and robustness of the effects (unresolved questions of external validity) and the mapping of particular ranges of IAT scores onto corresponding ranges of discriminatory behavior (unresolved questions of where, if anywhere, along the IAT scoring metric, discrimination arises against given groups).

The Need for Transparency and Communication

Because of its applied implications and controversial nature, we think that implicit-bias research needs an unapologetic dose of skepticism, supplemented with enforcement of Mertonian norms of data sharing and replication. This is good scientific advice in general, but it carries especially pointed relevance in this area, where researchers have moved quickly to apply their findings to public policy and legal disputes. In this spirit, we were transparent from the outset in our requests for data and about our project's purposes and we shared our analyses with both sets of authors (McConnell & Leibold, 2009; Ziegert & Hanges, 2009) before submission to this journal. Both researchers should be commended for participating in this project, when others stated that they either would not or could not comply with so basic a scientific norm as making their data available for reanalysis.

J. C. Ziegert and P. J. Hanges engaged in a dialogue with us prior to submission (personal communication, beginning March 3, 2006), which resulted in our conducting additional analyses in response to their statements. A. R. McConnell instead stated that he would only comment on our reanalysis during the formal review process (personal communication, March 28, 2007). That was his prerogative, but it resulted in a communication breakdown

Hart Blanton, Department of Psychology, Texas A&M University; James Jaccard, Department of Psychology, Florida International University; Jonathan Klick, School of Law, University of Pennsylvania; Barbara Mellers and Philip E. Tetlock, Haas School of Business, University of California, Berkeley; Gregory Mitchell, School of Law, University of Virginia.

Authorship order was determined alphabetically.

Correspondence concerning this article should be addressed to Gregory Mitchell, University of Virginia School of Law, 580 Massie Road, Charlottesville, VA 22903-1738. E-mail: greg_mitchell@virginia.edu

that we must describe in order to respond to McConnell and Leibold's (2009) assertion that our reanalysis focused on a variable named IATTRIM when it should have focused on a variable named IATFULL.

McConnell and Leibold (2009) state that it was "unclear why Blanton et al. would emphasize a different measure (i.e., the IATTRIM) than the one reported in the original article that they are critiquing" (p. 585) and suggest the correct variable was "in the data set provided to them" (p. 585). In fact, the IATFULL variable was not originally provided to us (McConnell only provided it to us during the review process, at our request), and the findings in McConnell and Leibold's (2001) study cannot be fully replicated using the IATFULL variable that they claim they used. We used the IATTRIM variable originally because McConnell sent us the IATTRIM variable and described it in his coding sheet as "the critical IAT score reported in the paper." In initial passes through the data set, we replicated key results using this variable (e.g., Table 2 in McConnell & Leibold, 2001), but we also found small discrepancies in some of the key analyses (particularly in Table 3 of McConnell & Leibold, 2001). We sought clarification on this matter when we sent McConnell an early draft of our reanalysis, but he indicated that he would only comment during the editorial review process. As a result of his taking that position, we could only make note of the discrepancy in our original submission to the journal.

Through later communications facilitated by the review process, we learned that a different coding of the IAT variable existed (IATFULL) that was not in the data set initially sent to us, and McConnell provided us the new IATFULL variable, indicating that we should be able to replicate all of McConnell and Leibold's (2001) results using this variable. That did not turn out to be the case. It appears that McConnell and Leibold (2001) used IATFULL for some of their analyses and IATTRIM for other analyses in their article, although there is no mention of this in the report. Because the means and standard deviations of the IATTRIM variable matched those reported in the article (Table 2) and because the distribution of this variable had properties suggesting to us that it was probably the product of data trimming (consistent with the Methods description in McConnell & Leibold, 2001) and because the IATFULL variable did not have these properties, we focused our reanalysis on this variable.

It was never our intention to focus attention on McConnell and Leibold's (2001) mistake in switching between IAT variables in their study because this switch produced relatively small differences in the results they reported. But the response by McConnell and Leibold (2009) necessitated a reply, so that readers understand that there was no error on our part and that there was nothing unclear about why we used the variable we did. We analyzed the data that McConnell provided to us following his instructions, focusing on a variable that replicated many of the key findings in the report, and McConnell is fully aware of the history of what transpired.

Our Other Variable Choices in McConnell and Leibold (2001) Were Appropriate as Well

McConnell and Leibold (2009) are concerned that we did not focus our analyses on all outcome variables that they studied. We focused on a variable that McConnell and Leibold (2001) identified in their original article as being the source of a primary outcome and toward which the majority of their analyses were directed. We also noted in our initial article, contrary to McConnell and Leibold's (2009) response, that the results of McConnell and Leibold (2001) were not uniform across outcomes, sometimes showing statistically significant associations with the IAT and sometimes not.

McConnell and Leibold's (2009) Table 1 is purported to be a complete listing of the results of the outcome variables in their original study, but it is not. A host of other outcome variables showing nonsignificant results are excluded from this table. We urge readers to examine the original report and form their own conclusions about how pervasive the IAT effects were on the many outcome variables reported, keeping in mind (a) the results of our reanalyses, (b) the demonstrated fragility of associations between the IAT and behavioral outcomes, and (c) the fact that no formal outlier analyses were undertaken in analyses for any of the outcome variables. The failure to conduct outlier analyses also applies to the McConnell and Leibold (2009) response, where McConnell and Leibold (2009) inappropriately use an outlier identified in one analysis as the single defining outlier in all analyses (see Table 1 in their response). We still contend that it is a large inferential leap to go from such results to statements about the necessity for major policy changes and the real-world legal implications of implicit prejudice in the United States.

Our Outlier Analyses Were Appropriate and Indicated a Need for Inferential Caution

We stand by our outlier analyses despite the questions raised in the responses. With respect to the age-based outlier in McConnell and Leibold (2001), we discussed the reasons why high IAT scores in older adults are suspect as measures of implicit bias, particularly when the original IAT scoring algorithm was used (as discussed by Greenwald, Nosek, & Banaji, 2003, the original IAT algorithm was likely to inflate bias scores for very young and older persons, which was one reason why they rejected further use of that algorithm; see also Hummert, Garstka, O'Brien, Greenwald, & Mellott, 2002). McConnell and Leibold (2009) question our focus on this older individual because they asserted that she did not have the largest IAT effect. But we never suggested that this person had the most extreme IAT score on the IATTRIM variable—and we provided the entire scatter plot of IAT scores and behavior to ensure transparent reporting. Moreover, in the new coding of the IAT variable that McConnell sent to us and argued we should use (IATFULL), this older participant did have the most extreme IAT effect. McConnell and Leibold (2009) fail to note this fact, and so this appears to be another case where they moved from emphasizing one coding of the IAT variable in one analysis to another coding in a different analysis. We would also note that the oldest individual produced the most extreme molar behavior rating in terms of anti-Black behavior as well.

Having said that, sheer extremity of a score on a single variable should not be the primary basis for outlier identification in correlational analyses and that was not our criterion in identifying this data point as an outlier. Outlier analysis is concerned with scenarios in which a single or a few data points mask more fundamental trends in the data and whether conclusions are robust to the presence or absence of the outlier(s). In the McConnell and Leibold (2001) data, significance patterns were indeed dependent on a single case, namely a 50-year-old woman in the midst of a more traditional group of college students.

Regarding Ziegert and Hanges's (2009) discussion of outliers, we agree that visual inspection of data may miss outliers. It was because of this that we applied robust statistical techniques for testing the difference in slopes in the two corporate conditions in their study. The robust methods generally are superior to the outlier-based strategies used by Ziegert and Hanges (2009) in their response (see Wilcox, 2005, for relevant discussions). It also is unclear whether Ziegert and Hanges (2009) conducted their outlier identification analyses (focusing on dfBetas, standardized residuals, and the like) using their original hierarchical linear modeling regression strategy or instead by using product terms in standard regression. If the latter, then the outlier methods they relied on are potentially problematic because of the behavior of products of scores (see McClelland & Judd, 1993). The test of slope differences using Theil-Senn regression that we used did not rely on product terms. It appropriately compared the simple bivariate slopes (based on Theil-Senn regression) in the two corporate climate conditions. Although the Theil-Senn estimation strategy has breakdown points (as Ziegert & Hanges, 2009, cite Wilcox to so indicate), it unquestionably has superior robustness properties to the regression methods relied on by Ziegert and Hanges (2009).

Regardless of whether the strategies suggested by us or our respondents are better, we hope everyone can agree that all outlier detection methods involve subjective judgments. As examples, statisticians disagree about the cutoff values that should be used to define outliers using centered leverage scores, dfBeta, and studentized residuals—despite the reference by Ziegert and Hanges (2009) to the somewhat arbitrary criteria discussed in Cohen, Cohen, West, and Aiken (2003). The inherent subjectivity of any outlier analyses suggests that researchers should be cautious about the use of strict rule-based strategies, such as those employed by Ziegert and Hanges (2009). In all cases, we reanalyzed data using robust methods of analyses and based our conclusions on those robust methods.

Ultimately, our point is not to debate outlier identification methods. Our point is that the significance patterns in these two studies are fragile and either hinge on the analytic method applied or on the presence of a single individual in the data set, so that one should avoid using the data to make strong statements about the pervasiveness of implicit prejudice in the American population. In our view, a researcher should be cautious about drawing strong conclusions or encouraging strong applications of an effect when key findings change with the removal of one or two people from a data set, no matter how those people are identified.

There Is More to Interjudge Reliability Than Correlations Between Judges

McConnell and Leibold (2009) are unconcerned that their results on the variable we focused on were driven by a single judge and did not hold up across judges. To us, this fragility indicates a result that should be interpreted with caution. Just as respondents can be outliers in a set of data, so too can a single judge be an outlier.

McConnell and Leibold (2009) criticize our decompositional analyses for the different judges by asserting that the interjudge correlations were strong, hence, such disaggregated analyses were unnecessary. We do not find interjudge reliability estimates in the 0.40 to 0.50 range to be strong, but that is for the reader to decide. In psychometrics, it is well known that examination of the simple correlation between two items (e.g., two judges) is a weak form of determining whether the items represent a common construct. In addition, the items should show similar correlations with external variables (proportional to their loadings on the construct). The low interjudge correlations and the fact that the judges' ratings showed differential significance patterns of correlations with the IAT should raise questions about their internal validity. We stand by our analyses and conclusions with respect to the judges.

IAT Zero Points Matter

McConnell and Leibold (2009) criticize our concern for the IAT zero point, but their argument ignores (often heated) debate on this very topic. Their response also places them in the odd position of espousing a view opposite to that promoted by the IAT architects (e.g., Greenwald, Nosek, & Sriram, 2006; Greenwald, Rudman, Nosek, & Zayas, 2006; cf. Blanton & Jaccard, 2006a, 2006b, 2006c, 2006d). The issue of the IAT zero point is critical because, if the race IAT does not have a rational zero point (such that those with positive scores have an attitudinal preference for Whites relative to Blacks and those with negative scores have an attitudinal preference for Blacks relative to Whites), then all of the IAT diagnoses offered through the Project Implicit Web site (https://implicit.harvard.edu/implicit/)—used to make inferences about the supposed prevalence of implicit racism—are flawed. This point was acknowledged by Greenwald, Nosek, and Sriram (2006), who argued that the rational zero-point assumption was warranted. Greenwald, Rudman, et al. (2006) went so far as to argue that researchers who use the IAT can apply unorthodox (and liberal) statistical tests—tests that normally would be off limits for social scientists. They can do this, they argued, because researchers can embrace the strong zero-point assumption if they work with the IAT (but not if they use self-report measures). It is thus problematic for McConnell and Leibold (2009) to state that "viewing IAT = 0 as some sort of meaningful criterion is at odds with the understood relative nature of the IAT" (p. 586).

The Evidence for a Link Between IAT Scores and Discrimination Is Weak

We disagree with McConnell and Leibold's (2009) claim that the extant literature unequivocally shows that the IAT or other implicit measures of attitudes predict meaningful discriminatory behaviors over and above explicit measures of attitude. First, there is a paucity of studies that seek to predict discriminatory behavior of any kind in real-world settings with meaningful conscious-attitude controls. Second, the issues one must address to draw strong conclusions are far more nuanced than McConnell and Leibold (2009) suggest, as we discussed in our original article.

We also draw attention to another recent reanalysis project that casts further doubt on the ostensible link between IAT scores and discriminatory behavior. Dawson and Arkes (2008) reanalyzed Green et al. (2007), which reported that "pro-White" IAT scores predicted anti-Black medical diagnoses in a sample of physicians. The reanalysis revealed that the physicians in this study, as an aggregate, actually showed a bias favoring Blacks relative to Whites, and this pro-Black bias became statistically nonsignificant among those participants with higher (pro-White) IAT scores. This finding is consistent with our reanalysis of the McConnell and Leibold (2001) data and further casts into doubt simple assumptions about the relation of IAT scores to racially discriminatory behavior.

Mundane Realism Has Its Place, But This Is Not It

Ziegert and Hanges's (2009) discussion of mundane realism misses the point of our discussion of the climate manipulation that they employed. We were not criticizing the particular manipulation they chose; we were simply noting that it places limitations on the generalizability of the data that cannot be removed by citing examples of deplorable behavior in the real world. Ziegert and Hanges's (2009) citation of Mitchell, Tetlock, Mellers, and Ordóñez's (1993) hypothetical-society studies actually bolsters our point: Further research by ourselves (Mitchell, Tetlock, Newman, & Lerner, 2003; Mitchell & Tetlock, 2006; Ordóñez & Mellers, 1993) and others (Michelbach, Scott, Matland, & Bornstein, 2003; Scott, Matland, Michelbach, & Bornstein, 2001) has shown that the findings of Mitchell et al. (1993) are context sensitive. We would not defend those who would use Mitchell et al. (1993) to argue for large-scale changes in societal policies bearing on income redistribution, and we are surprised that Ziegert and Hanges (2009) seem willing to embrace strong claims about changes to antidiscrimination law (such as in Kang & Banaji, 2006) based on their study, even assuming a mundane realism that simulates the effects of explicit racist directives within an organization.

As a related aside, we question Ziegert and Hanges's (2009) assertion that their study's materials—which explicitly directed college students to discriminate against Black employees in a fake hiring situation involving hypothetical applicants—somehow fall "in the middle" of a continuum of "mundane realism (i.e., the extent to which elements of the real world are incorporated in an experiment)" (p. 591). Ziegert and Hanges (2009) can define the endpoints of their hypothetical continuum as they wish, but their definitional maneuver should not disguise the fact that they have no real evidence of how frequent explicit racist directives of the type they used are in American workplaces in 2008, directives that would be illegal under existing law.

Ziegert and Hanges (2005) Found IAT Prediction Only When Bias Was Requested of Participants

The new partial-correlation analyses that Ziegert and Hanges (2009) report to assert IAT effects in both their experimental conditions are at odds with conclusions stated in their original article (Ziegert & Hanges, 2005) and uses a model that is statistically misspecified and, hence, inappropriate. Originally they hypothesized that a linear trend would be observed between IAT scores and ratings in the racial bias climate but not in the equality climate, and they claimed to find this pattern. Now, in their response, they imply that the trend appeared in both conditions because the correlation between the IAT and the criterion remains statistically significant after controlling for the (dummy scored) manipulation (but ignoring the interaction effect, i.e., omitting the product term). Neither their original theory and results, nor the simple main-effects tests we ran, suggest that this new description of their data is correct. In fact, if an interaction is present as they assert in their original article and in their reanalysis, it is not appropriate to conduct the partial correlation analysis they did to make this new conclusion because it applies a misspecified model to the data (i.e., it omits an interaction effect that is supposedly present). Such specification error can bias parameter estimates, bias standard errors, and renders significance tests ambiguous. This new analysis can distract readers from our points of agreement with Ziegert and Hanges (2005) regarding links between their theory and their results.

Ziegert and Hanges (2005) Pursued a Novel IAT Coding

Ziegert and Hanges (2009) assert that their scoring of the IAT is consistent with the approach taken in the latest IAT scoring algorithm (Greenwald et al., 2003). Ziegert and Hanges (2005) is the only published study to rely on the scoring method that they chose, rendering that method unique by definition. We have seen no data on the comparability of their algorithm and the new scoring algorithm, and Greenwald et al. (2003) indicated no interest in incorporating error rates into the IAT score in their evaluation of the different methods, contrary to the statements of Ziegert and Hanges (2009). Given the sensitivity of IAT scores to scoring method, it was appropriate to point out this unique feature of Ziegert and Hanges's (2005) experiment and note that its impact on the interpretation of their results is unclear.

No Consistent Pattern of Preference Emerged Across the Two Studies

Ziegert and Hanges's (2009) defense of the pattern of candidate ratings in their study reinforces our criticism of McConnell and Leibold (2001). We agree with Ziegert and Hanges (2009) that stronger associations between IAT scores and treatment of the Black candidates in their study seems consistent with theory. But, the opposite pattern was found in McConnell and Leibold (2001; and see Dawson and Arkes, 2008). Our point was that no consistent pattern is emerging in the IAT literature on discrimination, and this reality was obscured because these two research groups both oriented their analyses around inspection of a difference-score criterion, one that was open to multiple interpretations.

Concern With the IAT Scores and Behavior of Individuals

The focus by Ziegert and Hanges (2009) on expected values for prediction intervals is not appropriate to the question we asked in our original article. Our question was whether one can say with confidence—based on their data—whether a given person with a given IAT score would show discriminatory behavior. One cannot. A different question a researcher might ask (and the question asked by Ziegert and Hanges, 2009, in their prediction interval analysis) is whether the regression equation generated from the data yields predicted, aggregated group means on the outcome variable that can confidently be said to be above zero for a given IAT profile. This question is different than the one we posed. Our conclusion stands with respect to the prediction intervals: Knowing the IAT score for any individual in Ziegert and Hanges (2005) study, no matter how extreme that score is in the observed distribution, does not allow one to state with reasonable confidence whether that individual will or will not show discrimination on the outcome variable. Ziegert and Hanges (2009) conclude from their analyses that “the vast majority of our participants had prediction intervals that did not overlap with zero” (p. 595), but this interpretation is misleading because their intervals are not focused on individuals. The intervals in their analyses speak to predicted means for groups of individuals, not any given individual.

We are in complete agreement with Ziegert and Hanges’s (2009) argument that factors other than predictive utility need to be taken into account to determine the practical significance of a test or a score on that test, a point we also have emphasized in our past work (e.g., Blanton & Jaccard, 2006a, 2006b). This fact makes it that much more difficult to make any substantive statements about the practical utility of the IAT and what different scores on it might mean in applied contexts, which was one of the points of our initial article.

Greater Attention Is Needed to Explicit Constructs

Ziegert and Hanges (2009) performed analyses that purportedly show that the IAT predicts behavior over and above two global measures of explicit racist attitudes. We do not find these analyses convincing because modern attitude theory long ago rejected the use of the global attitude measures they used to predict specific behavioral criteria. Ziegert and Hanges (2009) are using an outdated model of attitudes that no modern day attitude theorist would embrace. A more reasonable approach, for example, would compare the predictability of implicit attitudes over the constructs represented by Ajzen and Fishbein’s theory of reasoned action (see Ajzen, 1991). To date, investigations using such strong explicit measures are not to be found in the IAT literature.

In Closing

McConnell and Leibold (2009) close their response by endorsing replication but arguing that trust is the cornerstone of scientific openness. We agree that replication is crucial to scientific progress but disagree with conditioning replication on trust. If researchers are allowed to avoid sharing data by questioning the motives behind the data reanalysis, then it becomes all too easy to seal large bodies of empirical work from critical scrutiny. Too much trust among ideologically like-minded investigators can lead all too easily into scientific groupthink, and too much deference to other researchers in order to gain trust and avoid conflict will lead to scientific stagnation rather than progress.

In our view, this exchange highlights the need for more rigorous scrutiny of the predictive utility of the race IAT. Millions of people have been told that their IAT scores indicate a weak, moderate, or strong implicit bias against African Americans, with the express suggestion that such bias may result in discriminatory behavior. Prominent social scientists and law professors are now using these data to argue for court action and changes in public policy to remedy an alleged epidemic of hidden prejudicial tendencies. Yet not a single empirical study has been conducted in a way that allows researchers to identify the range of IAT scores that reliably results in discriminatory behavior. Moreover, although it has been argued that positive (or anti-Black) scores on the race IAT reflect implicit prejudice against Blacks, our reanalysis of McConnell and Leibold (2001) as well as our reanalysis of other data sets by other researchers (Dawson and Arkes, 2008) show aggregate trends in which respondents with seemingly anti-Black IAT scores act more positively toward Blacks than Whites, if not equally so. In addition, our analyses of prediction intervals suggest that the relationship between the IAT and behavior is so fragile that one cannot state with confidence that any given individual will perform discriminatory behavior, no matter what score he or she received on an IAT. No theoretical, methodological, or analytical posturing changes these facts. Would the psychology profession be so accepting if psychiatric assessments were handled with the casualness with which some social scientists label respondents racially biased?

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50,* 179–211.

Blanton, H., & Jaccard, J. (2006a). Arbitrary metrics in psychology. *American Psychologist, 61,* 27–41.

Blanton, H., & Jaccard, J. (2006b). Arbitrary metrics redux. *American Psychologist, 61,* 62–71.

Blanton, H., & Jaccard, J. (2006c). Postscript: Perspectives on the Reply by Greenwald, Rudman, Nosek, and Zayas (2006). *Psychological Review, 113,* 166–169.

Blanton, H., & Jaccard, J. (2006d). Tests of multiplicative models in psychology: A case study using the unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 113,* 155–165.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Dawson, N. V., & Arkes, H. R. (2008). Letter to the editor: Implicit bias among physicians. *Journal of General Internal Medicine, 24,* 137–140.

Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I., & Banaji, M. R. (2007). The presence of implicit bias in physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of General Internal Medicine, 22,* 1231–1238.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197–216.

Greenwald, A. G., Nosek, B. A., & Sriram, N. (2006). Consequential validity of the Implicit Association Test: Comment on the article by Blanton and Jaccard. *American Psychologist, 61,* 56–61.

Greenwald, A. G., Rudman, L. A., Nosek, B. A., & Zayas, V. (2006). Why so little faith? A reply to Blanton and Jaccard’s (2006) skeptical view of testing pure multiplicative theories. *Psychological Review, 113,* 170–180.

Hummert, M. L., Garstka, T. A., O’Brien, L. T., Greenwald, A. G., & Mellott, D. S. (2002). Using the Implicit Association Test to measure age differences in implicit social cognitions. *Psychology and Aging, 17,* 482–495.

Kang, J., & Banaji, M. R. (2006). Fair measures: A behavioral realist revision of “affirmative action.” *California Law Review, 94,* 1063–1118.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114,* 376–390.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37,* 435–442.

McConnell, A. R., & Leibold, J. M. (2009). Weak criticisms and selective evidence: Reply to Blanton et al. *Journal of Applied Psychology.*

Michelbach, P. A., Scott, J. T., Matland, R. E., & Bornstein, B. H. (2003). Doing Rawls justice: An experimental study of income distribution norms. *American Journal of Political Science, 47,* 523–539.

Mitchell, G., & Tetlock, P. E. (2006). An empirical inquiry into the relation of corrective justice to distributive justice. *Journal of Empirical Legal Studies, 3,* 421–466.

Mitchell, G., Tetlock, P. E., Mellers, B. A., & Ordóñez, L. D. (1993). Judgments of social justice: Compromises between equality and efficiency. *Journal of Personality and Social Psychology, 65,* 629–639.

Mitchell, G., Tetlock, P. E., Newman, D. G., & Lerner, J. S. (2003). Experiments behind the veil: Structural influences on judgments of social justice. *Political Psychology, 24,* 519–547.

Ordóñez, L., & Mellers, B. A. (1993). Tradeoffs in fairness and preference judgments. In B. A. Mellers & J. Baron, J. (Eds.), *Psychological perspectives on justice: Theory and applications* (pp. 138–154). New York: Cambridge University Press.

Scott, J. T., Matland, R. E., Michelbach, P. A., & Bornstein, B. H. (2001). Just deserts: An experimental study of distributive justice norms. *American Journal of Political Science, 45,* 749–767.

Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd ed.). San Diego, CA: Academic Press.

Ziegert, J. C., & Hanges, P. J. (2005). Employment discrimination: The role of implicit attitudes, motivation, and a climate for racial bias. *Journal of Applied Psychology, 90,* 553–562.

Ziegert, J. C., & Hanges, P. J. (2009). Strong rebuttal for weak criticisms. *Journal of Applied Psychology.*

Received September 3, 2008
Revision received October 14, 2008
Accepted October 23, 2008 ■