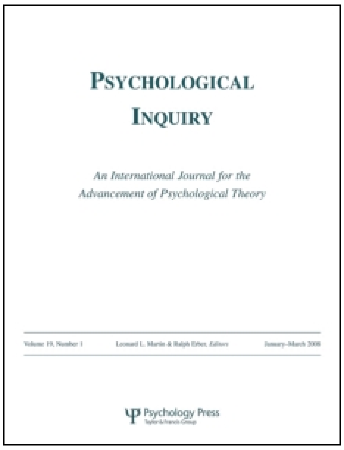

Psychological Inquiry
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775648164

AUTHORS' RESPONSE: The Implicit Prejudice Exchange: Islands of Consensus in a Sea of Controversy

Philip E. Tetlock [a]; Hal R. Arkes [b]
[a] Haas School of Business, University of California, Berkeley. [b] Department of Psychology, Center for Health Outcomes, Policy, and Evaluation Studies, Ohio State University.

Online publication date: 19 November 2009

To cite this Article Tetlock, Philip E. and Arkes, Hal R.(2004) 'AUTHORS' RESPONSE: The Implicit Prejudice Exchange: Islands of Consensus in a Sea of Controversy', Psychological Inquiry, 15: 4, 311 — 321

To link to this Article: DOI: 10.1207/s15327965pli1504_03

URL: http://dx.doi.org/10.1207/s15327965pli1504_03

Psychological Inquiry
2004, Vol. 15, No. 4, 311–321

AUTHORS' RESPONSE

The Implicit Prejudice Exchange: Islands of Consensus in a Sea of Controversy

Philip E. Tetlock
Haas School of Business
University of California, Berkeley

Hal R. Arkes
Department of Psychology
Center for Health Outcomes, Policy, and Evaluation Studies
Ohio State University

Some commentators dismiss Arkes and Tetlock (this issue) on two grounds that we regard as specious, namely that we are either (a) political apologists for covert bigotry—the soft-on-racism charge raised by Sears (this issue) or (b) psychological naïfs who cling to obsolete definitions of attitudes and prejudice—the scientific incompetence charge raised by both Banaji, Nosek, and Greenwald (this issue) and Wittenbrink (this issue). To move the debate beyond such caricatures, the scholarly community must eventually confront the specific empirical objections that Arkes and Tetlock raised about how construct-validational work is done (whether researchers are giving fair weight to ideologically dissonant alternative explanations) and the specific logical objections that Arkes and Tetlock raise about the blurring of fact-value distinctions (whether psychological theory is advanced by researchers' opining on the political defensibility of their research participants' opinions).

Responses to our target article fall into three categories: those who see our contribution as a dangerous distraction from the serious business of advancing empirical knowledge of prejudice (Banaji et al., this issue; Sears, this issue), those who agree with us on some points but disagree on others (von Hippel, this issue; Wittenbrink, this issue), and those who mostly agree with us and even aggressively expand key parts of our argument (Redding, this issue; Suedfeld, this issue).

We devote the bulk of this reply to rebutting the five most serious charges leveled against us, specifically that (a) we are impeding the search for ever more subtle and insidious forms of prejudice, (b) we are wedded to a ridiculously narrow conception of attitudes and ignorant of the scientific advances over the last two decades that reveal the power of unconscious attitudes, (c) we harbor a "quaint" conception of prejudice and have failed to keep pace with the evolving disciplinary consensus on what constitutes prejudice, (d) we have offered aid and comfort to racial profilers and other miscellaneous scoundrels, and (e) we endorse a doctrine of guilt by racial association. We do not dwell, however, exclusively on points of disagreement. We note at several junctures points of convergence among contributors. And we stress the potential of adversarial collaboration and Bayesian reputational bets to challenge the more dubious assumptions underlying implicit prejudice research and to open an ideologically tendentious research program to the recognition that reasonable people can attach a wide range of meanings to terms like *prejudice* and *racism*.

The "Why Are We Making It Harder to Identify Covert Racism?" Charge

No matter how impressive the scientific credentials of implicit prejudice researchers, and those credentials are deservedly widely esteemed, their opinions about what constitutes a prejudiced opinion—about where we, as a research community, should set our thresholds for labeling thoughts, feelings, or actions as prejudiced or racist—deserve no special deference. For prejudice is a *value* as well as a *factual* judgment. And the problem is, at root, not just a psychological one. It is a deeply political one that requires us to make moral judgments of which aspects of public opinion (implicit or explicit) deserve our censure—and make no mistake here, when you call people prejudiced, they feel censured.

Imagine a continuum. At one end are the most demanding standards for labeling thought or action as racist. One must believe that a particular ethnic-racial group is genetically inferior to one's own group, endorse antimiscegenation laws, and support de jure segregation. Abraham Lincoln would have qualified as a racist by these standards (Donald, 1996), but over the last 50 years it has become increasingly difficult to find citizens with this attitudinal profile. Many social scientists have responded to this powerful historical trend by lowering their thresholds for labeling citizens as racist. They are convinced that, although the norms regulating public debate on race have changed dramatically, underlying racial attitudes have not changed nearly as much: Racism has gone underground and social scientists need to forge sophisticated new methodological tools for detecting these covert or symbolic or modern or implicit forms of racism.

But how far down the continuum should we venture in pursuit of ever sneakier forms of racism? And on whom should the burden of proof fall: those who claim to have uncovered hidden reservoirs of racism in public opinion or those who uphold the traditional jurisprudential and scientific standard of "not guilty (or no effect) until demonstrated beyond a reasonable doubt?" We believe serious accusations require serious scrutiny—and the more radical the indictment, the heavier the burden of proof the prosecution (or promoters of a theory) should bear.

It proves disturbingly difficult, however, to achieve agreement on who should shoulder the burden of proof and on how close research programs come to satisfying those burdens. In 1986, a sharp disagreement—one to which Sears (this issue) alludes—arose. Sniderman and Tetlock (1986a, 1986b) wrote a pair of papers critical of symbolic racism theory. According to this theory, racism in America in the 1960s and 1970s underwent a transformation. Far from disappearing, it had assumed new, subtle forms that allowed it to continue operating as a potent force in American politics. This new racism was theoretically anchored in strong moral feelings that African Americans collectively violate traditional American values such as individualism, the work ethic, and even obedience and discipline. In this framework, opposition to affirmative action testified to symbolic racism. So too did opposition to busing. In fact, racism and opposition to busing in the original formulation were literally the same thing: Symbolic racism was operationally defined as opposition to busing and affirmative action.

Tetlock (Sniderman & Tetlock, 1986a, 1986b) was convinced back in 1986, and remains convinced, that work on symbolic racism did not come close to satisfying even minimal burden-of-proof standards. He saw the strategy of equating conservative policy preferences with racism as a political and scientific dead-end: A strategy that, by politicizing the concept of prejudice, forecloses a scientific answer to the hypersensitive question of how big a role racial hostility plays in driving conservative preferences.

Tetlock (Sniderman & Tetlock, 1986a, 1986b) had not, however, appreciated in 1986 how methodologically and theoretically creative the pursuit of covert prejudice would become—and how low the thresholds for making attributions of racism would fall. The research that Arkes and Tetlock (this issue) review is not grounded in tendentious tautologies but rather in state-of-the-art, cognitive-neuroscience models of human memory. And, it is not necessary for people to endorse any beliefs or attitudes whatsoever to be labeled as implicitly prejudiced. Implicit prejudice is calibrated in millisec of response-time facilitation or inhibition and the inference of racial animus cross-validated by eyeblinks and brain scans. Arkes and Tetlock question a far more scientifically formidable research program than Sniderman and Tetlock (1986a) did.

But Sears (this issue) insists on stressing the similarities between the 1986 and this exchange. He characterizes the target article for this journal and the earlier contribution by Sniderman and Tetlock as "both… critical of new lines of research describing an underrecognized but powerful form of racism in the post-civil rights era." And he portrays both articles in the same negative light: politically motivated critics offer glib armchair critiques of the hard, honest work of empirical researchers dedicated to uncovering new subtle forms of prejudice.

We reject his characterization. But, to explain why, there is no avoiding another dollop of armchair philosophizing. If we want to stop talking past each other, we need to start clarifying what we mean when we use certain words. *Racism* is not a value- or ideology-neutral descriptive term of the sort one would expect to encounter in the data language of a neo-positivist science. *Racism* is not on an epistemological par with *spreading activation networks* or *response time latency* or *amygdala activation*. Charges of racism carry powerful connotative as well as denotative judgments: Such charges imply that we, the investigators, are condemning as well as describing the attitude in question. Racism is a political hot potato that partisans are quick to disavow for themselves and to attribute to their adversaries (Tetlock, 1994).

Arkes and Tetlock (this issue) observe—and no one disputes—that the meaning one attaches to racism hinges on where one places oneself in the roiling debates over racial inequality in early 21st-century America. Liberals tend to set their thresholds of proof lower than do conservatives, and social psychologists tend to be liberal. If we insist on importing value-laden constructs from political debate, like racism, into our descriptions of scientific work, we should expect political fissures to ripple through the scientific debate. We should expect those on the left to be more concerned with underestimating the tenacity of prejudice, less concerned with exaggerating it, and more inclined to answer *yes* to questions like the following: Should we cross-validate our implicit measures against explicit ones that define racism in terms of opposition to busing or affirmative action? Against measures that are sensitive to preferences for White versus African-American roommates in college? Against measures of greater eyeblinking or discomfort in presence of African-American versus White confederates?

It is, in our view, profoundly misleading to treat racism as though it were a well-defined property of human psychological functioning awaiting the right measurement breakthrough. We subscribe to an old-fashioned philosophy of science that distinguishes between factual and value judgments, between *is* statements, such as "reaction time fell from 3,290 to 3,260 millisec," which can be true or false "nonperspectivally," and *ought* statements, such as "someone with an associative network with these properties should be censured for racism," which can only be true or false "perspectivally." The *ought* claims are perspectival for essentially the same reason David Hume offered two centuries ago for the logical irreducibility of *ought* propositions to *is* propositions: We can justify such claims only by importing moral assumptions into the debate that give us grounds, extrascientific grounds, for judging courses of action as, to varying degrees, laudable or deplorable. There is inevitably a logical disjunction between factual observations (*x* possesses property *y* with probability *z*) and the moral-political evaluations we attach to those observations (we approve or disapprove of *y*). To express the argument in another way, there is no inconsistency in applauding the impressive "factual" discoveries that psychologists have made about the correlates of implicit attitudinal measures but balking at giving any special weight to the attitudes these psychologists have toward the implicit attitudes of their research participants.

Early 20th-century behaviorists prudently respected this fact–value distinction (Boring, 1950), but early 21th-century psychologists have ignored it. As a result, much social psychological work on implicit prejudice now rests on a classic category mistake, a logical confusion over the types of predicates that scientists can justifiably assign to their objects of study. The deepest irony of this exchange is that it places some of social psychology's greatest experimentalists (e.g., Anthony Greenwald, Russell Fazio) in the same camp as social psychology's most prominent postmodernist, Kenneth Gergen. A sophisticated experimental effort to ground social psychological work on prejudice securely in basic memory and neurological processes, thereby reinforcing the discipline's status as a positive science, yields as its end product claims about racism that can only be "true" in historical context, conditional on the political norms prevailing at particular moments in particular subcultures.

Interestingly, Banaji et al (this issue) grant part of this argument. They are acutely aware that the disciplinary and wider societal norms for determining who is prejudiced have evolved. Indeed, they take us to task for our "old fashioned" mid-20th-century, Gordon-Allport-style views. But they see the emergence of new, relaxed norms for labeling attitudes as prejudiced as a sign of scientific progress, not a change in political fashion reflecting new power alignments in the broader society. We doubt we could disagree more deeply. We are, moreover, baffled by the "scientific" argument that they offer for decoupling prejudice and antipathy. Put simply, to avoid an extensive side tracking of our argument, the coexistence of widespread positive attitudes toward women and, say, salary inequities is not the slam dunk demonstration they imply it to be. A large and econometrically sophisticated literature has been devoted to trying to tease apart the complex causation underlying sex differences in compensation. Banaji et al.'s treatment of this literature is, in our view, a further worrisome sign of their readiness to pronounce prejudice in the midst of profoundly ambiguous evidence.

Is there an escape from this morass? The radical solution is almost Skinnerian in its severity: Abandon all pretense of political relevance and retreat into a less value-laden data language in which we replace loose talk about *prejudice* and *racism*, that almost everyone understands (or, more accurately, thinks they understand) with a more precisely descriptive set of terms, such as the *dimensionality of negative affectivity*, that only the cognoscenti understand. This prescription flows from the fact–value distinction central to the neo-positivist principles underlying experimental psychology (Boring, 1950), but it is unlikely to have broad appeal among social psychologists, many of whom would view it as an abdication of their moral responsibility to stay silent when they are convinced their data speak directly to deeply divisive issues.

The more moderate solution would be to engage in adversarial collaboration, for researchers representing different theoretical camps to agree, ex ante, on what would constitute fair empirical tests of their respective positions (in effect, Bayesian reputational bets) and to update their beliefs in accord with the results (cf. Mellers, Hertwig, & Kahneman, 2001). To jumpstart this process, it would, of course, be critical to reframe the debate, to move from an all-or-none dichotomy in which one either is or is not sensitive to the continuing threat of racism to a matter-of-degree continuum along which reasonable people can disagree. It would also be critical to clarify both where and why different camps set their thresholds for making politically charged attributions (in effect, accusations) of prejudice and racism. But, once we work through these awkward preliminaries of articulating our own political prejudices about racial prejudice, we could begin to capture the debate in a Bayesian reputational-bet format.

We suspect that the first discovery will be that implicit racism researchers approach this topic with relatively strong prior hypotheses that covert or unconscious racism in early 21st-century America is a powerful force whereas Arkes and Tetlock (this issue) approach with more skeptical priors. We suspect the second discovery will be that different camps have markedly different views about the diagnosticity of various types of evidence—be they eyeblinks or amygdala activation or modern racism scales. Here we need to move beyond the impressionistic assessments of construct validity that often prevail in social psychology to much finer grained assessments of the conditions under which particular lines of evidence gain or lose probative value. For instance, when should Arkes and Tetlock abandon their shame-and-guilt counter-interpretation of eyeblink data? Or when should implicit prejudice researchers give more weight to the alternative, generally more benign, explanations advanced by Arkes and Tetlock? We suspect the third discovery will justify the entire laborious process: We will find out who is willing to honor the ex ante reputational bets they made about the likelihood of different empirical outcomes conditional on their favorite hypotheses about the causal mechanisms driving human responses to implicit associative measures.

We do not view this proposal as a panacea. There will be occasions when the feuding camps balk at recognizing the diagnosticity of each other's data. One particularly tough test for our adversarial-collaboration/Bayesian-bet proposal arises when we explore the theoretical implications of possible patterns of correlations between implicit and explicit attitudinal measures. We see no point in wading deeply into this messy literature: Correlations have ranged widely (Fazio & Olson et al., 2003) and there seem to be no theoretical constraints to prevent researchers from adopting a "heads-I- win-tails-I-do-not-lose" stance in which they can claim that positive correlations suggest convergent validity, that zero order correlations suggest discriminant validity and that negative correlations even suggest repression or overcompensation. But Banaji et al. (this issue) believe that recent results offer support to their view of what the Implicit Association Test (IAT) measures: After subtracting out unreliability, the correlations between the IAT and certain explicit measures of prejudice are both positive and statistically significant.

Assuming this is true, what conclusions should we draw? Here it is critical to stick as close to the facts, and as far from the researchers' value-charged abstractions, as possible. If the explicit measure is the Modern Racism Scale (and it often is), we will have learned that there is an association between the tendency for college students to think African Americans collectively should rely more on themselves and less on government and the tendency for students to have more difficulty responding quickly to evaluatively dissonant word pairings in the IAT. Does this mean that the slower-reaction-time respondents to the IAT are more likely to be racists—as Banaji etl al. (this issue) apparently do? Or does it, among other possibilities, simply mean that these respondents to the IAT are more likely to believe—perhaps correctly, perhaps not—that the primary obstacles to racial inequality are now internal to the African-American community, a position endorsed by many White and African-American pundits, not all conservative?

Our view—the one we would have entered into a Bayesian reputational bet—is that the data are insufficiently diagnostic with respect to the question of whether the IAT is tapping implicit prejudice. And we see the confidence of Banaji et al. (this issue) in the diagnosticity of the data as an ominous sign of how badly the field needs to implement our model of adversarial collaboration. But, we should not get off the hook too easily. It is fair to ask: What would induce us to change our minds? The answer is simple: Show us strong associations between the IAT and measures of prejudice toward African Americans that pass the three Allportian tests (antipathy, rigidity, and erroneous belief) and we will be willing to make some Bayesian belief updating adjustments of our own. Of all the empirical claims that Banaji et al. make, the following reference to an unpublished manuscript struck us as most promising for making truly high-diagnosticity reputational bets: "A recent meta-analysis by Poehlman, [Uhlmann, Greenwald & Banaji] (2004) [shows] that implicit attitudes not only predict, but that they predict better than explicit measures when the target measure is social group discrimination."

Such evidence would indeed be potentially problematic for our position. But note that we say *potentially* because—given Banaji et al.'s (this issue) expansive conception of prejudice—we worry that they may be working with a comparably expansive conception of discrimination. Let us therefore stipulate that if the term *discrimination* in the Banaji et al. claim meets the classic legal standard of treating otherwise equally qualified Black and White human beings differently for jobs, housing, or admission to school, we would indeed count this as evidence against an important part of our position, suggesting as it does that implicit prejudice measures can be validated even against old-fashioned Allportian criteria, in which there is a prospect of serious harm being inflicted on a group simply by virtue of its skin color, a claim that would be taken seriously in a court of law.

Having put our cards on the table, however, it is only fair to ask: What pattern of associations between implicit and explicit measures of prejudice would induce Banaji et al. (this issue) to change their minds? We see a pattern in the literature: When implicit measures predict explicit indicators of prejudice, investigators take such findings as construct-validational support but when they fail to predict such indicators, they take the findings as evidence of how skillfully some or even most respondents are at cognitively overriding their unconscious prejudicial impulses. Even more troubling, the best individual-difference moderator measure of the tendency to engage in override—the motivation to control prejudice scale (Dunton & Fazio, 1997)—does not allow us to escape this closed-to-disconfirming-evidence loop. For, this carefully constructed scale is itself open to alternative interpretations. *Prejudice* is—it should be clear by now—a term with ever-shifting political meanings. Are high scorers on the scale reining in raw bigotry or are they displaying hypersensitivity to even matter-of-fact racial differences—exactly the sorts of hypersensitivity that would arise if they feared that esteemed professors would label them as racist if they displayed even trace awareness of unpleasant facts of life in our society? The distinction matters.

Of course, even if we can reach agreement on the ex ante terms for reputational bets, there is still no guarantee of ex post convergence (just greater normative pressure to converge than there otherwise would be). Indeed, if past work on cognitive conservatism and belief perseverance in political judgment is any guide (Tetlock, in press), we should expect widespread reluctance to changing one's mind even when thoroughly unexpected results materialize. Although we can never reduce each side's wiggle room for escaping disconfirmation to zero (reputational bets can never anticipate every contingency), we can agree that, insofar as one side needs to resort much more frequently than the other to "rewriting" their ex ante reputational bets ex post, that side should rethink its assumptions about causal mechanisms—and be taken less seriously if it fails to do so.

We could dismiss all this as cheap talk—and David Sears (this issue) does. But we view it as establishing the conceptual preconditions for scientific progress. Banaji et al. (this issue) are wrong to imply that we ignored the construct-validational evidence for implicit measures of prejudice. We just have different perspectives on the probative value of their evidence. We also disagree with Banaji et al. if they believe that meta-analysis is a substitute for conceptual analysis of competing philosophical and political perspectives on what constitutes prejudice. There are many thoughtful observers outside this highly specialized program of research who—as we shall see shortly—find the Banaji et al. conception of prejudice far too expansive—so expansive that it runs the risk of trivializing a political-psychological force responsible for the death of many millions in the last century.

The "Are We Psychologically Naïve?" Charge

Another way to cut the debate short is to argue that Arkes and Tetlock (this issue) take positions that are so psychologically extreme as to be laughable. We see signs of this strategy in both the Banaji et al (this issue) and Wittenbrink (this issue) critiques. Both critiques take us to task for embracing far too narrow and old-fashioned a conception of attitudes, and both go on to make compelling cases for more expansive conceptions of attitudes that make room for both implicit affective associations and consciously justified beliefs. Unfortunately, both critiques rest on the incorrect assumptions that Arkes and Tetlock believe implicit measures cannot capture attitudes and that attitudes must always be exclusively defined as sentiments that people explicitly and consciously endorse. And both critiques compound the misunderstanding by stipulating that, if we accept their more reasonable definition of attitude, we are also logically locked into (by implication) accepting a definition of prejudice consistent with current patterns of usage in the implicit prejudice literature. Indeed, Banaji et al. interpret our reluctance to accept their definition of prejudice as a sign that Arkes and Tetlock subscribe to the "notion that genuine prejudice is only consciously reportable prejudice."

We do not endorse that notion either. Our actual psychological position on some of these issues is surprisingly close to our toughest critics in some respects. We have maintained all along that people often harbor evaluative dispositions that they are unwilling or unable to articulate. Indeed, we declare in the target article that "implicit attitudes are plausible psychological constructs that serve plausible psychological functions." We also continue to applaud the efforts of several contributors to this exchange—Banaji, Nosek, Greenwald, and Wittenbrink—to shed light on the antecedents and consequences of these implicit attitudes (even if we see no value added, and considerable value subtracted, when researchers superimpose their own value judgments on these attitudes). And we continue to feel that it would be silly to straight-jacket ourselves to the point where we can only identify those bigots who publicly endorse views that pass virtually all political thresholds for bigotry.

We did, however, raise the possibility—and recognize we are far from the first to do so—that implicit associative measures are substantially influenced by stereotypes that people know are widely held but that they may not in any sense endorse. As Judd, Blair, and Chapleau (2004) observed in their ingenious study designed to sort out why African-American, relative to White, faces automatically facilitate the categorization of handguns, "It is one thing to believe that the tendency to misclassify weapons in the hands of young African Americans is part of a negative, prejudicial bias on the part of police officers. It is a rather different thing to believe that the officers are influenced by a widely shared stereotype of African Americans as both violent and athletic, unreliable and possessing a good sense of humor, lazy and street-smart" (p. 77). The two different assessments imply different views of American society as well as different remedies.

We also raised the possibility in the parable of the two Jesses that such implicit associative measures capture types of negative affect (e.g., shame, guilt, embarrassment, annoyance, fear) that should not be confused with those that are normally associated with racial hostility, but that could be so confused because they produce many hypothesized consequences of racial hostility, such as gaze aversion and conversational bumbling. Related but distinct, we raised the interpersonal awkwardness hypothesis that the common variance between implicit associative measures and certain construct-validational criteria, such as seating or roommate preferences, may be produced by concern that Blacks have not yet collectively forgiven Whites for past mistreatment. Causality is tricky here but it matters whether Whites are avoiding Blacks because they dislike Blacks or because they fear that Blacks dislike them. These often-neglected counter-interpretations have implications for reputational bets: They undercut, in our view, the diagnosticity of much of the evidence invoked by implicit prejudice theorists.

And we raised the possibility that some fractional component of the remaining negative affect captured by implicit measures is traceable to realistic assessments of group differences that are responsive to new evidence, rather than to sweeping prejudgments that are impervious to nuance or distinctions.

For all these reasons we argue that scoring "badly" on implicit measures is not automatically indicative of racial animus and that indeed the current construct-validational evidence is nowhere close to justifying such an inferential leap.

Finally, we noted our discomfort with holding people morally—perhaps some day legally—accountable for unconscious affective associations that, if drawn to their attention, they might well disavow. If one is going to charge an individual with implicit racism, then the preconditions for guilt should be present. There is an inherent tension between the psychological level of analysis at which implicit prejudice researchers work (extremely rapid unconscious associations) and the moral-political debate that they wish to enter (in which issues of legal liability loom large; the doctrine of *mens rea*). We believe that neither the IAT by itself nor in concert with ambiguous construct-validational criteria (the Modern Racism Scale, eyeblinks, or amygdala activation) is sufficient to sustain charges that rise to the level of moral and legal culpability. Indeed, our sense is that indictments for charging implicit racism take us precariously close to the realm of thought crimes —a point we tried to make by asking how the research community might have reacted to McCarthyite measurement efforts to identify implicit Marxists in the 1950s or would react now to efforts by Attorney General Ashcroft to identify implicit Islamo-fascists in the first decade of the 21st century.

Insofar as we take serious psychological issue with Wittenbrink (this issue) and Banaji et al. (this issue) over the ethical status of implicit attitudes, it is over two points: (a) their hypothetico-deductive certainty that if we accept that implicit measures tap into racial attitudes, we are thus obliged to accept that those measures also tap into prejudice; and (b) their optimistic faith that if we only recognized that research is a slow self-correcting process, we would be more impressed by how far work on implicit prejudice has advanced and less inclined to complain about how inconclusive the evidence to date is.

We do not accept the implied automatic equivalence of negative racial attitudes and racism. We resist in part because we are old-fashioned Humeans (as well as Allportians) who still honor the, admittedly sometimes fuzzy, fact–value distinction. Negative affectivity is close to a purely factual or *is* statement whereas racism is a prototypic value-laden *ought* statement—a statement that manifestly means different things in different social worlds at different times.

We also resist the implied equivalence of negative racial attitudes and racism because we are reluctant to indict people for negative affectivity grounded in awareness of depressingly real covariations between group membership and attributes normally suffused with negative affect, be those attributes liberals prefer to talk about (e.g., a history of slavery) or attributes conservatives prefer to talk about (e.g., current patterns of criminality, poor performance in school or family breakdown). In the spirit of signal detection theory, we recognize that where observers place their thresholds for making value-laden attributions of racism can be reasonably influenced by objective base rates of threatening behavior as well as, of course, the moral value they place on avoiding Type I errors (failing to identify a real threat) and Type II errors (false alarming). *Rational racism* and *Bayesian bigots* strike us as oxymorons.

We wish we could join in Wittenbrink's (this issue) and Banaji et al's (this issue) optimism that the self-correcting scientific method is unfolding as it should in the domain of implicit prejudice. There are, to be sure, some promising signs: At least one prominent researcher has retreated from his early bold claims that implicit measures provide a "bona fide pipeline" to true racial attitudes (Fazio, Jackson, Dunton, & Williams, 1995). But we see the need for a forceful contrarian nudge to encourage researchers to give alternative explanations a fair shake and to recognize there is much more interpretive ambiguity shrouding widely used construct-validational measures than commonly supposed in the introductory or discussion sections of many articles. And that, of course, is why we wrote our critique in the first place.

The "Failure to Keep Up With Changing Fashions" Charge

Banaji et al (this issue) refer to Gordon Allport as simply a scientist of his time who could not cope with the paradigm-busting insight that one could be prejudiced even if one's attitudes were not laden with rigid, ungrounded-in-reality, antipathy toward outgroups. We agree that historical context is important: But Allport (1954), unlike the authors, had an excuse for his "quaint" views: His life was drawing to a close in the mid-20th-century, and his quaint conception of prejudice was organized not around social slips and slights but rather around the horrors of concentration camps and the humiliations of de jure segregation.

This historical digression brings us to yet another misconception underlying the Banaji et al. (this issue) argument. For them, prejudice is the epistemological equivalent of a rock that scientists can subject to physical tests to determine its structure and origins. Banaji et al. act as though social psychologists have discovered over the last 50 years that the rock of prejudice comes in many previously unexpected forms: It need not have antipathy or rigidity or be organized around incorrect beliefs. And that is what makes Arkes and Tetlock's (this issue) ignorance so inexcusable: It is our failure to keep up with the rapidly evolving corpus of knowledge about the novel psychological properties that the rock of prejudice possesses. This position strikes us as untenable. Prejudice does not have an objective existence external to the research community. Prejudice is a fuzzy-set moral category. There is wide agreement on the prototypes, genocidal violence and lynching, and it becomes increasingly difficult to achieve political consensus as one moves outward toward the periphery. It is not a matter of researchers discovering the novel properties of prejudice. It is a matter of researchers discovering that they can get away with making value judgments about other people's opinions that would once have been derided—even by the most liberal opinion—as presumptuous.

Banaji et al. (this issue) see no downside risk in defining prejudice down or in mingling factual and value judgments. They are convinced there is an inexorable historical trend in their favor, toward defining prejudice down, and that our objections have an antiquarian-racist, Plessy-versus-Ferguson flavor. But political forecasting is a notoriously imprecise science (Tetlock, in press), and prophecies have a nasty habit of boomeranging back on the prophets. Lenin was wrong: The communists, not the capitalists, were consigned to the ash heap of history. What if Banaji et al. are wrong? Do we want a social psychology as a science to take moral-political stands on what attitudes citizens are permitted to hold without being exhorted to eternal vigilance or consigned to thought reform? What if, by some strange twist of history, the "wrong people" wind up in charge of the commanding heights of academia: the journals, and the funding agencies, and the tenure committees?

Our concerns on these scores are reinforced by the commentaries of Redding (this issue) and Suedfeld (this issue). They too observe that implicit prejudice research sets the threshold for convicting citizens of racism far too low, all the way down to the level of "thought crime." We also agree with Redding and Suedfeld that there is a cost to a scientific discipline of adopting a partisan perspective on prejudice, anointing it as a scientific consensus, and then challenging the political motives of those who question the wisdom of this scientific strategy. And we confess to sharing Redding's and Suedfeld's bemusement that it is we who stand accused of politicizing a hitherto apolitical research program on implicit prejudice. It apparently falls to us to point out that, well before this exchange hit the presses, implicit prejudice research was already controversial and had provoked sharp rebuttals in the elite press.

Perhaps we social psychologists should not care what conservatives think of our research. Science should not be swayed by outside political pressures. We agree with the principle (it is in the spirit of the fact–value distinction) but worry about its implementation in this context. Is it not possible, given the ideological lopsidedness of our field, that our science has already been swayed by both internal and external political pressures to lower its thresholds for attributions of prejudice to unprecedented lows? And is it bad manners even to raise the possibility that social psychologists have "politicized" themselves by treating terms of political abuse as explanatory constructs? Consider this characterization of implicit-prejudice research by Stein (2004) from *The Wall Street Journal*:

> To discern the hidden presence of racial bias in white thinking, the researchers confronted a group of …

white undergraduates with images and words associated with black people. … Then, … they measured activity in the student's brains. It turned out that the higher the test subjects scored on the bias meter …, the worse they performed on a test of cognitive ability. The effect of contact with another race taxed the biased white brain to the point of making it unable to concentrate. Imagine the excitement of the researchers as "objective" research confirmed what they had known all along: that Jim Crow might be dead and buried but that residual racism survives in parts of the white brain—even if unconnected in any way to actual behavior.

Yet, … knowing who these kids were, is it not just as likely that all the test showed was their hyper-sensitivity not to be racist … ? Perhaps true, study author Jennifer Richeson replies, but struggling too hard not to appear racist shows that one is uncomfortable with members of other races—ergo, still biased. … Talk about defining bias down! (p. 16)

Some social psychologists find it easy to dismiss Stein's (2004) protestations as defensive conservative ranting. But it is not so easy to dismiss the research results that von Hippel (this issue) discusses in that part of his rejoinder in which he takes us to task for excessive timidity. He notes that Frantz, Cuddy, Burnett, Ray, and Hart (in press) have shown that White students find the IAT threatening if they were told that it was a measure of prejudice or if they ascertained that fact themselves as they took the test (a connection that is not difficult to make). In such cases, stereotype threat emerges from participants' concern that they will be perceived as prejudiced because they are White. This threat disrupts performance on the IAT, much as it does with other groups and other performance domains, by causing people to provide even larger IAT effects than they otherwise would.

Even if one does not care about hurt feelings among Whites falsely accused of racism, one might care about the impact on the African-American community. Redding (this issue) cites the work of Gilbert (1998) and others who found that African-American college students often made erroneous attributions of prejudice based on negative feedback that they had received in ambiguous social situations. Such misattributions of prejudice can be as corrosive for race relations as are prejudicial attitudes themselves. And, in this vein, recall from our target article that Sniderman (2003) has shown rather hefty correlations between African Americans' perceptions that Whites are prejudiced and African Americans' endorsement of conspiracy theories that attribute ills confronting the African-American community—from AIDS to illegal drugs—to sinister external influences. Exaggerating the tenacity of prejudice can be as lethal an error as understating it.

The "Are We Trying to Justify Racist Conduct?" Charge

von Hippel (this issue) takes us to task for excessive boldness when we suggest it may be *rational*—as that term is conventionally defined among specialists in judgment and decision making—for people (a) to use racial base rates that imply unflattering comparisons for traditionally disadvantaged groups, (b) to use the discounting principle when doing so causes us to discount the abilities of affirmative action beneficiaries, and (c) to act like expected utility maximizers when doing so requires responding differently to people of different races. von Hippel makes a valiant effort in each case to reconcile academic norms of political correctness and social science norms of rationality. But the blunt fact is that neither we nor he possess the necessary data. We can ascertain the predictive power of race in his various street scenes only by doing real-world assessments of ecological-cue validities. We can ascertain the logical defensibility of discounting the abilities of minority versus athletic versus legacy admits to universities only by first gauging the magnitude of the advantage conferred by each status. It would not bother us one iota if von Hippel proved closer to the mark on all these real-world assessments. What matters is that we agree to treat these sensitive issues as empirical uncertainties rather than simple assume that anyone who uses racial base rates or discounts for affirmative action advantages is, ipso facto, racist. In the spirit of Sears's (this issue) plea to avoid premature closure, we ask only for suspension of judgment pending better evidence.

Both von Hippel (this issue) and Redding (this issue) do however raise a useful flag of caution to those too eager to embrace our purely hypothetical arguments about the rationality of using racial base rates. What is expected utility maximizing for the individual may be decidedly suboptimal for society at large. Both commentators also point to the danger of self-fulfilling prophesies and, in doing so, capture the essential tragedy of the evolving dilemma of race in American life. Heeding base rates is logical. As long as the races differ statistically on criminality and other socio-economic outcomes to which most Americans attach evaluative significance, those who heed base rates will hire fewer African Americans, grant loans to fewer African Americans, and so on. To prevent vast numbers of innocent people from being injured by individual Whites assimilating individual African Americans into unflattering generalizations, we, as citizens, think it was a good idea—in custom and in law—to turn racial base rates into "forbidden base rates" (Tetlock, Kristel, Elson, Green, & Lerner 2000). But this is a public-goods argument that, as game theorists well know, has no bearing on issues of individual rationality. As social scientists, our mission is to understand the world, not to change it by acting as norm enforcers who stigmatize as racist anyone who displays the slightest trace of even unconscious understanding of patterns of covariation that almost all of us wish would vanish.

Finally, von Hippel (this issue) correctly notes that "although disliking an outgroup is clearly more prejudicial than simply liking one's own group more, there is a long tradition of treating relative preferences as prejudicial." We can probably agree here that the differences between us are a matter of degree. Suedfeld (this issue) notes the enormous changes that have occurred in American responses to social-distance scales over the last 70 years. Whereas the average score for African Americans once hovered between granting them citizenship and allowing them only as visitors, the average score now hovers between acceptance as a best friend and acceptance as suitable marriage partner for members of one's most intimate group. We have indeed come a long way from Selma Alabama (not to mention Auschwitz) if these social-distance data capture the downward trajectory of prejudice in American society. The clarifying question becomes: How small would relative differences need to be, and how effusive would the absolute level of acceptance need to become, before researchers should entertain the possibility that American public opinion has changed dramatically for the better over the last two generations. As Suedfeld notes:

> In a world that has seen, comparatively recently, the massacres of Armenians by Turks, Jews by the citizens of most European nations, Tibetans by the Chinese, Tutsis by Hutus, Bosnians by Serbs and vice versa, and—at a lower numbers, but equally lethally—ethnic or religious murders all over the world, it seems strange to obsess about racism as revealed only by slower reaction times or brain activation.

But Banaji et al. (this issue) find nothing strange in what Suedfeld (this issue) views as a *reductio ad absurdum* scenario. In fact, Banaji et al. raise von Hippel (this issue) in the moral competition over who is more vigilant against prejudice (note the ominous similarities to the group polarization literature). Banaji et al. are prepared to dispense with all three components of Allport's (1954) definition of prejudice: "an antipathy based on a faulty and inflexible generalization" (p. 9). They require neither antipathy nor even differential levels of positive affect. They require neither that the generalization be faulty nor even that it be unresponsive to evidence. For them, an attitude is suspect if it leads to disparate impact.

This admission strikes us a startling. Under Banaji et al's (this issue) definitional regime, we are justified in labeling as prejudiced any mindset that is, in field or experimental designs, linked to departures from equality of result. Banaji et al. thus embrace what Arkes and Tetlock (this issue) view as another *reductio ad absurdum* of their position. We found it disturbing that implicit prejudice researchers define prejudice so low that anyone who lives in a society with inequalities that have evaluative significance is virtually guaranteed to score as prejudiced. Banaji et al. do not find it disturbing. But we still wonder: Why blame the psychological messenger? If Banaji et al. have zero tolerance for all forms of intergroup inequality, why dress up their political preference as psychological theory? Scientific nomenclature is supposed to be precise. Why not call prejudice what it has become: those attitudes, implicit or explicit, among research participants that provoke the requisite degree of disapproval from researchers of a certain political persuasion? And scientific theory is supposed to be parsimonious: Why not explain exactly how our psychological understanding of attitudes is deepened by researchers' editorializing on the moral and political defensibility of those attitudes? And if one cannot, then apply Occam's razor to the superfluous posturing.

Some Stray Misconceptions

Banaji et al. (this issue) correctly assert that modern views of justice proscribe punishing those who are only nominally associated with a guilty party. But Banaji et al. erroneously assert that a consequence of our viewpoint is the legitimization of the doctrine of guilt by association. They reach this conclusion by blurring, once again, the fact–value distinction. Our point was a purely factual one: People who notice a true covariation between race and some characteristic will tend to associate the two factors. This simple principle, as old as psychology itself, explains Rev. Jackson's fear that the pedestrian behind him is African American, notwithstanding the overwhelming majority of African-American pedestrians intend Jackson no harm. We never offered the value judgment that this was a Good Thing. We only suggested that implicit prejudice researchers suspend their value judgments and acknowledge that Jackson (and others in like circumstances) are not necessarily bigots.

We also did not say that people, like Rev. Jackson, engage in complex expected utility calculations when making decisions in street scenes of this sort. And we most emphatically did not say that people engage in calculations similar to those listed by Banaji et al. (this issue) in their Table 1! Banaji et al. reach this erroneous conclusion by confusing facts and values again (this time, taking a normative conclusion about good decision processes and transforming it into a factual one about how people characteristically think).

For the record, we simply stipulated that if a person processes options in a manner congruent with the prescriptions of expected utility theory, it is difficult to deem that person irrational or otherwise cognitively impaired. All that Rev. Jackson or any pedestrian needs to know is that African Americans and Whites have different crime rates, and the conclusions from our expected utility example will be valid. No complicated calculations are required. The calculations were done by us to show that Jackson's behavior comported with expected utility theory. Needless to say, we do not think that he crunches the numbers.

Closing Thoughts

Where does this exchange leave us? The purely logical parts of our argument emerge unscathed. There is no serious challenge to the following points:

1. The discounting principle is a sign of good judgment in nonracial domains but is a sign of prejudice in racial ones.
2. The use of base rates is a sign of good judgment in the judgment-and-decision-making literature but a sign of bigotry in the prejudice literature.
3. The maximization of expected utility is the benchmark of rationality in most arenas but is a sign of bigotry in the racial arena.
4. Fast-and-frugal detection of cue covariations is lauded by Gigerenzer and Goldstein (1999) as a marvel of human cognition, but it is a source of sinister motivations according to many social psychologists.

The more empirical parts of our argument remain more open to challenge, hinging as they do on how observers of varying psychological and political persuasions interpret the construct-validity evidence for implicit measures. Observers with different causal assumptions and value priorities—unless anchored down by ex ante reputational bets—will always be able to put their preferred theoretical spins on the same facts. Consider a final illustration of the potential for partisan mischief by researchers with low thresholds for ferreting out covert racism. On September 6, 2003, Mayor Coleman of Columbus, Ohio, who is African American, announced he would increase police presence in two areas of the city, both heavily African American. Is he as bad as L.A. Police Chief Parker, whose motives Sears (this issue) questions? Is this an example of modern racism? Or is this a rational response? Let's presume that Mayor Coleman was blinking a lot during this announcement and his amygdala was firing at an abnormally high rate. Would this physiological data combined with his announcement make him a bigot? Because his announcement cannot be clearly defined as race-sensitive or race-insensitive in modern America, it is hard to know what to make of that amygdala activity. And because Mayor Coleman obviously associated African-American neighborhoods with criminality more than he associated White neighborhoods with criminality, he is fated to look bad on implicit associative measures.

In closing, there is no point disguising the depth of our disagreement with some of our critics. Many implicit prejudice researchers see themselves as embarked on a valuable societal as well as scientific mission. And history may vindicate them. Arkes and Tetlock (this issue) may be hopelessly out of step with the times: early 19th-century Humeans and mid-20th-century Allportians, who cannot keep up with the fast pace of scientific breakthroughs on implicit prejudice in the early 21st century. We may be the "old" and "false"; the implicit prejudice researchers the "new" and "true"—off to the museum with us.

But we are skeptical, so skeptical that this exchange reminds us of a famous epigram of an even more famous political theorist, Karl Marx, who observed that history repeats itself, the first time as tragedy, the second time as farce. We cite this epigram not in any way to derogate our critics but rather to crystallize the biggest of all issues on which we and they diverge. We all agree that race relations in the United States have long been laced with tragedy, but we seem alone in worrying that the dialectic is now propelling us into a new phase populated by African-American taxi cab drivers who stand accused of racism for their reluctance to pick up certain types of passengers in certain neighborhoods, a city employee fired for using the Scandinavian-based word *niggardly*, college professors who are condemned for failing to support affirmative action admissions quotas, and, manifestly the most absurd of all, by dog lovers who level charges of "canine racism" against measures to curb attacks by pit bulls in New York City (Schauer, 2003). Banaji et al. (this issue) are, in this sense, correct in characterizing us as anti-zeitgeist. We suspect that, when the history of social psychology is written at the end of the 21st century, implicit prejudice research will be a prime exhibit of how society became so obsessed with avoiding stereotypes that it skewered citizens as racists for displaying even trace awareness of politically painful realities. We are even prepared to make that a reputational bet—although either collecting or paying presupposes extremely optimistic projections of our life expectancies.

Note

Philip E. Tetlock, Haas School of Business, University of California–Berkeley, Berkeley, CA 94720-1900. E-mail: tetlock@haas.berkeley.edu. Hal R. Arkes, Department of Psychology, Ohio State University, 240N Lazenby Hall, 1827 Neil Avenue, Columbus, OH 43201. E-mail: arkes.1@osu.edu

References

Boring, E. G. (1950). *A history of experimental psychology*. New York: Appleton.

Donald, D. H. (1996). *Lincoln*. New York: Simon & Schuster.

Dunton, B. C., & Fazio, R. H. (1997). An individual-difference measue of motivation to control prejudice. *Personality and Social Psychology Bulletin, 23,* 316–326.

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

Fazio, R., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobrrusive measure of racial attitudes: A bona fide pipline? *Journal of Personality and Social Psychology, 69,* 1013–1027.

Frantz, C. M., Cuddy, A. J. C., Burnett, M., Ray, H., & Hart, A. (in press). A threat in the computer: The race Implicit Association Test as a stereotype threat experience. *Personality and Social Psychology Bulletin*.

Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: Take the best heuristic. In G. Gigerenzer, P. M. Todd, & The ABC Research Group, *Simple heuristics that make us smart.* New York: Oxford University Press.

Gilbert, D. J. (1998). The prejudice perception assessment scale: Measuring stigma vulnerability among African American students at predominately Euro-American universities. *Journal of Black Psychology, 24,* 305–321.

Judd, C. M., Blair, I. & Chapleau, K. (2004). Automatic stereotypes vs. automatic prejudice: Sorting out the possibilities in the Payne (2001) weapon paradigm. *Journal of Experimental Social Psychology, 40,* 75–81.

Mellers, B. A., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science, 12,* 269–275.

Schauer, F. (2003). *Profiles, probabilities, and stereotypes*. Cambridge, MA: Belknap Press of Harvard University Press.

Sniderman, P. (2003). *Black pride and Black prejudice*. Princeton, NJ: Princeton University Press.

Sniderman, P. M., & Tetlock, P. E. (1986a). Reflections on American racism. *Journal of Social Issues, 42,* 173–188.

Sniderman, P. M., & Tetlock, P. E. (1986b). Symbolic racism: Problems of motive attribution in political debate. *Journal of Social Issues, 42,* 129–150.

Stein, H. (2004, January 30). Dumb and dumber. *Wall Street Journal*, p. A15.

Tetlock, P. E. (1994). Political psychology or politicized psychology: Is the road to scientific hell paved with good intentions? *Political Psychology, 15,* 509–529.

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78,* 853–870.

Tetlock, P. E. (in press). *Expert political judgment: How good is it? How can we know?* Princeton, NJ: Princeton University Press.