Preferences with Social Constraints: Political Correctness and Decision-Making

Michael I. Norton               Joseph A. Vandello and Andrew Biga
Massachusetts Institute of Technology       University of South Florida

John M. Darley
Princeton University

Manuscript under review, *Organizational Behavior and Human Decision Processes*
Abstract

Across a range of social contexts (e.g., employment decisions, college admissions), individuals are often faced with the unenviable task of selecting among people of different races, decisions which are problematic because they raise the specter of racial bias. We propose that people – particularly White Americans – are uncomfortable making such decisions and will go to great creative lengths to justify them. This research demonstrates people’s unwillingness to cite race as a factor in decision-making, then shows how the use of racial information impacts information-processing. In a first study, Whites who hired a Black marketing representative for a hypothetical clothing manufacturer were unwilling to acknowledge race as a factor, even when explicitly licensed to do so. In a second study, participants overwhelmingly chose Black candidates over equally qualified White candidates for admission to college, but justified the decision using arbitrary non-racial criteria; White participants carried these criteria forward in a subsequent task.

Key words: Stereotypes, prejudice, discrimination, political correctness, decision making, attribute importance, justification, distortion
Preferences with Social Constraints: Political Correctness and Decision-Making

In the wake of a Fifth Circuit Court of Appeals decision calling into question the use of race as a criterion for admission to colleges and universities (Hopwood v. Texas, 1996), admissions officers in Texas who were motivated to favor minority applicants for admission to college – whether due to their own egalitarian motives, to meet desired quotas, or simply to appear politically correct – were faced with a difficult challenge: How could applicants be favored on the basis of their race when the use of such information was prohibited? More generally, when people are motivated to use forbidden information (such as race or gender) in their decisions, how do they go about doing so without appearing biased? The New York Times listened in on an admissions committee at Rice University in Texas which was faced with this dilemma, and found they managed to favor such applicants by searching through their resumes for other, nonracial information which would support that candidate’s admission (Steinberg, 2002).

Decision-makers are frequently under pressure – both self-imposed and from others – to justify their behavior, especially when this behavior might be seen as biased. Indeed, social pressures to justify one’s decisions can have a powerful impact on people’s preferences (e.g., Simonson, 1989; Tetlock, 1992). Consider those scenarios with which we opened this paper – a college admissions officer is motivated to favor a Black candidate over a White candidate for admission to college. In one sense this is a classic binary choice paradigm between two options with varying features, though in general decisions like these are ones which individuals tend to solve in quasi-rational ways at best (e.g., Highhouse, 1997). In another, more social sense, this decision is quite different – social norms regarding race and the treatment of racial issues make this decision more loaded than many choices. Because such decisions face increased scrutiny, the
admissions officer may be reluctant to cite race as a reason for her decision. Now imagine that it so happens that the Black candidate has a higher GPA, but has taken fewer Advanced Placement (AP) classes, while the White candidate has a lower GPA but has taken more AP classes. In the present research, we show how individuals motivated to favor candidates based on race will deny that this factor influenced their decision, instead citing the importance of other attributes as justification: In this example, our admissions officer would admit the Black candidate by claiming that she simply saw GPA as a more important qualification than number of AP classes. This paradigm allows us to go further, however, in that we can vary the attributes that favor the preferred candidate: In some conditions, the qualifications are reversed, such that the Black candidate has the lower GPA but more AP classes, while the White candidate has the higher GPA but fewer AP classes. Across the two tasks, we can show that while people are clearly selecting based on race, they will flexibly inflate the value of whichever attribute happens to favor the Black candidate (see Hodson, Dovidio, & Gaertner, 2002; Norton, Vandello, & Darley, in press). Thus while their post-decisional attribute weightings appear to be consistent with their choice, the fact that GPA is seen as more important in one condition (when the Black candidate has the higher GPA) and AP classes are seen as more important in the other (again, when the Black candidate happens to have more) shows us that this seeming consistency is due far more to their desire to mask their decisions than to their prior beliefs about the importance of these qualifications.

*Justifying and Rationalizing Decisions*

In general, of course, people are motivated to justify their choices, both to themselves and others, especially when their actions might be perceived negatively. When acting in self-interested ways, for example, people are motivated to continue to see themselves positively – and
to invent excuses for such behavior (Batson, Thompson, & Chen, 2002; Monin & Norton, 2003); this desire to see the self in positive and preferred ways has been demonstrated in many domains across many situations (e.g., Kruglanski, 1989; Kunda, 1990; Pyszczynski & Greenberg, 1987). But people’s ability to engage in questionable behavior is limited by the extent to which they feel it can be justified. In Hsee (1996), for example, people underappraised the value of an apartment their fiancé was interested in purchasing – engaging in questionable behavior – but only when the true worth of the apartment was ambiguous. Similarly, Snyder, Kleck, Strenta, and Mentzer (1979) showed that participants who were given a choice of watching a movie with either a handicapped or non-handicapped person only sat with the non-handicapped person when the two targets were watching different movies; when the two targets were watching the same movie, participants sat with the handicapped person. In both Hsee (1996) and Snyder et al. (1979), people expressed bias only when there was a plausible alternative justification. One means of accomplishing the goal of being biased is to mask the true reason for one’s decision by justifying the decision in more acceptable terms. In these studies, we explore these issues in the domain of race, capitalizing on the fact that people are uncomfortable choosing between individuals based on race, and are motivated to justify their behavior when they do so. More specifically, we show that people who engage in questionable behavior – favoring individuals for jobs or college admissions on the basis of their race – will mask the reason for this decision by recruiting more acceptable attributes of the individuals – such as their academic performance – to justify that decision. We show that this casuistry – specious reasoning in the service of justifying questionable behavior – is not merely a strategy designed for public consumption, but also causes people to truly change their weights of attributes, and most importantly that these shifts in weights perpetuate through time.
The fact that people’s preferences and valuations are malleable – and sometimes easily influenced – is well-established (e.g., Ariely, Loewenstein, & Prelec, 2003; Payne, Bettman, & Johnson, 1992; Slovic, 1995). At the same time, however, people do seek to provide coherent reasons for their preferences (Nisbett & Wilson, 1977; Shafir, Simonson, & Tversky, 1993). While these reasons may not always reflect the true reasons for people’s preferences, individuals’ search for justification for their preferences can have interesting effects. It is well established empirically, for example, that people interpret – and distort – information to support their preferences, both prior to a choice (e.g., Brownstein, 2003), as choices are being made (e.g., Russo, Medvec, & Meloy, 1996; Russo, Meloy & Medvec, 1998; Simon, Krawcyzk, & Holyoak, 2004) and subsequent to choice (e.g., Brehm, 1956; Knox & Inkster, 1968; Mather, Shafir & Johnson, 2000); indeed, one of the key predictions of dissonance theory (Festinger, 1957) is that individuals seek to gain consistency by emphasizing the positives and de-emphasizing the negative aspects of their choices. One clear means of doing so, one that has been investigated particularly in the domain of binary choice, is to view information that favors one’s preferred choice more positively, and information that favors the rejected alternative more negatively. Many of the studies in this domain of research provide participants with two options, each with relevant information, and then track the distortion of information as a favorite begins to emerge (e.g., Russo et al., 1998), demonstrating how information can be distorted when individuals are trying to form preferences; given the need to be consistent, it is perhaps not surprising that people distort information. The current studies explore preference formation when people are trying to hide their preferences – or at least couch the reasons for their preferences in other terms.

Finally, the present investigation explores the consequences of the information distortion that accompanies the formation and justification of preferences, investigating the extent to which
such preferences perpetuate through time. When people are motivated to hide the true reasons for their preferences – and shift their attribute weights of other features to do so – we show that these shifted attribute weights can then impact subsequent choices, suggesting that hiding preferences in one decision can lead to real change in preferences on others.

**Norms of Political Correctness**

While there are many domains in which people are motivated to mask their decision-making rationales, issues of racial preference are an area in which this motivation may be particularly strong. White Americans are increasingly uncomfortable with issues of race (Swim & Miller, 1999), and while it is difficult to truly know whether privately held attitudes towards Black Americans have changed for the better, it is certainly true that the expression of negative attitudes in public life has decreased markedly over the last 50 years, due in part to the development of norms of political correctness. This discomfort with race may be justified, as there are real costs to violating norms of political correctness. Accusations of racism can sidetrack careers (as with Senator Trent Lott or the baseball pitcher John Rocker), and racists are among the most socially undesirable groups in America (Crandall, Eshleman, & O’Brien, 2002; Sommers & Norton, 2004). At the same time, those who follow norms of political correctness have been accused of pro-minority bias, or “reverse discrimination,” as with lawsuits over the legality of affirmative action in college admissions. While pressures have developed to favor minorities in certain settings to the extent that many Whites are quite willing to do so (as with affirmative action initiatives) it can still be problematic to admit that the minority status of an applicant played a role.

Because of strong norms against bias (whether it be anti- or pro-Black), the using of race as a factor in decision-making is essentially proscribed in many decision-making scenarios.
Nonetheless, decades of social psychological research demonstrate that social categories such as race play a substantial role in impression formation, person perception, and decision-making (see Fiske, 1998). Norms of political correctness dictate two somewhat curious behaviors: preference for minority candidates in domains such as employment or college admissions, but at the same time a prohibition on citing racial factors as the sole justification for that decision. Even if one favors a minority candidate for a position for purely egalitarian reasons, it is essential that the decision be justified in other terms – again, to avoid backlash like that witnessed in the debates over affirmative action in college admissions. In general, theories of modern or aversive racism suggest that individuals are only willing to engage in biased behavior to the extent that they are able to rationalize and justify it (e.g., Gaertner & Dovidio, 1986; McConahay, 1986). Though these theories applied to anti-Black prejudice, the taboo against using race as a factor in decision-making applies whether one is biased against minorities (as when one is racist) or biased in favor of minorities (due to a desire to be or appear egalitarian, or as a result of one’s support for affirmative action). In the former case, citing race as a factor is forbidden by law, while in the latter admitting that one is favoring a minority solely on the basis of his/her race requires acknowledging that you took the person’s race into consideration, a problematic behavior in a culture that stresses colorblindness (e.g., Plaut, 2002; Pollock, 2004). Even those individuals committed to affirmative action for egalitarian reasons may be unwilling to cite race as a reason for favoring a minority, since this kind of acknowledgment can undermine this person’s credibility in the workplace (e.g., Heilman, Block, & Stathatos, 1997). In sum, while individuals can be motivated to favor minorities – and alter their preferences to do so – they can be unwilling to admit the true reason for their preferences.
Interestingly, fear of violating norms of political correctness may be limited to members of majority groups – such as White Americans – who are particularly concerned about being evaluated negatively by members of minority groups (Vorauer, Main, & O’Connell, 1998); members of minority groups may have “credentials” that allow them to use racial information differently than members of majority groups (see Monin & Miller, 2001), a possibility we explore in Study 2.

Overview

In the studies reported in this paper, we first demonstrate individuals’ extreme reluctance to cite race as a factor in decision-making – even when the use of such information is explicitly licensed – in a simulated employee selection task in which participants are instructed to favor a Black candidate for a job (Study 1). In Study 2, using a college admissions task, we show that participants will favor Black candidates for admission to college over equally qualified White candidates, but refuse to admit that race was a factor, instead citing other qualifications of the candidates as their justification. In addition, we show that reweighting of attributes to favor Black candidates impacts subsequent decisions between candidates as well, suggesting the long-term impact of casuistry, though these effects were limited to White and not minority participants.

STUDY 1: Denying the Use of Race

In the first study, we wanted to stack the deck in favor of using race, to show that even when race is explicitly a factor in a decision, and moreover even when a situation warrants mentioning race as factor, the taboo still holds against any mention of social category as a basis for judgment. One way in which individuals feel they can violate norms of political correctness by using racial information to make decisions is if they are not held personally accountable – for
example if an authority figure instructs them to use race as a criteria (Brief, Dietz, Cohen, Pugh, & Vaslow, 2000). While such instructions may legitimate the use of race in decision-making, racial justifications for such decisions are still problematic. We thus expected that even when participants were licensed to use race in decisions, they would stop short of explaining their decisions in racial terms.

Method

Seventy-five White Princeton undergraduates (34 males, 41 females) completed the experiment for course credit. They read a short scenario in which they were told to imagine they were a mid-level manager of a clothing company in charge of hiring a new marketing representative. All participants were then told that a substantial percentage of the company's client base was African-American; half the participants were told their supervisor instructed them to hire an African-American as a result, while the other half were told that they themselves chose to do so. A brief description of the candidate they hired followed, including a picture of the candidate and information about his previous work experience and family life. Participants were then told that they received an email from either the African-American candidate or their supervisor asking them how they arrived at their decision, and were asked to write an email response. We thus varied both source of the decision (self vs. supervisor) and the target of the email (Black employee vs. supervisor).

We coded participants' response emails for whether they mentioned race as a criterion in their decision. We would clearly expect those who chose the candidate themselves to be hesitant about admitting using race as a criteria; we also expect, however, that even though instructions from authority figures can legitimize biased decision-making (Brief et al., 2000), such instructions fail to legitimize biased explanations for those decisions. Even when participants are
given instructions from a legitimate authority figure, then, we expect that their need to justify their use of race will make them reluctant to cite race as a factor in their decision-making. We also expected that while participants would be very unlikely overall to cite race as a factor in their decision, they would be even less likely to do when emailing the Black applicant himself, as such an email might be seen as inappropriate and might undermine that person’s confidence and status in the workplace (Heilman et al., 1997).

To further explore the public nature of this unwillingness to violate norms of political correctness, we also explored how participants’ own reported concerns with prejudice impacted their justifications, by having participants complete Plant and Devine’s (1998) measure of motivation to respond without prejudice, which has two subscales: the IMS measures participants’ internal motivation to avoid being prejudiced (e.g., “I attempt to act in nonprejudiced ways toward Black people because it is personally important to me”), while the EMS measures participants’ external motivation to avoid appearing prejudiced (e.g., “I attempt to appear nonprejudiced toward Black people in order to avoid disapproval from others”). Because selection decisions such as these – and the explanations that must accompany them – are impacted by public norms of political correctness, we expected participants’ willingness to cite race to be negatively associated with their scores on the EMS; because norms of political correctness have their strongest effects on public behavior, we did not expect that scores on the IMS would be related to mentions of race.

**Results and Discussion**

Two independent raters coded emails for use of race as a justification for the hiring (i.e., coding for use of words such as “Black,” “African-American” or “race”). Inter-rater reliability was quite high (Cronbach’s $\alpha = .96$).
Overall, only 10% of participants cited race in their emails, $\chi^2 (1, N = 75) = 49.61, p < .001$, though this factor was the explicit reason for the hire. As predicted, participants were even less likely to mention race when writing to the candidate (3%, 1/38) than writing to their supervisor (16%, 6/37), $\chi^2 (1, N = 75) = 4.09, p < .05$. Participants were no more likely to mention race whether they chose the candidate (11%, 4/37) or were instructed to do so (8%, 3/38), $\chi^2 < 1, ns$. Even when participants were licensed to use race by a supervisor’s demand, and even when then explaining that decision to the same supervisor, race was rarely mentioned.

We were concerned that one alternative explanation for why so few people mentioned emails in the responses to their supervisor was that rather than being uncomfortable mentioning race participants may have thought it unnecessary to mention this face because it was assumed, given the supervisor’s explicit instruction to hire a minority. To address this alternative explanation, we added an additional condition ($N = 15$) in which participants were instructed by their supervisor to hire a candidate from the Philadelphia area, then justify that decision to the supervisor. In contrast to the 16% who used race in emails to supervisors above, fully 80% of these participants mentioned the candidate’s Philadelphia connection in their return email, $\chi^2 (1, N = 15) = 5.40, p < .05$.

We next explored the relationship between participants’ individual differences in their concern with being prejudiced. As expected, participants’ reluctance to use race was related to their concern with appearing prejudiced to others (the EMS), $r = -.23, p < .05$: Participants more concerned with appearing egalitarian were less likely to cite race as a factor. Participants’ internal motivation to avoid prejudice (IMS) was not correlated with mentions of race, $r = .12, p = .30$. 
In sum, while business justifications such as instructions from a legitimate authority figure (Brief et al., 2000) can legitimize decision-making based on race, those who make such decisions may still be highly reluctant to admit using race, demonstrating people’s strong preference for justifying decisions – even those clearly made on the basis of race, and even those made in favor of a minority applicant – in non-racial terms. Of course, participants in Study 1 did not make actual decisions, but were instructed to select a candidate on the basis of race. Study 2 forces participants both to make decisions on the basis of race and then justify them.

**STUDY 2: Forming and Perpetuating Biased Preferences**

Study 2 was designed to show that people use race in decision-making when desired, but engage in casuistry to justify those decisions; in this case, justifying biased college admission selections by creatively reshaping candidates’ credentials. We created a hypothetical college admissions selection task in which participants were forced to choose between White and Black candidates for admission to college. We predicted that participants would be drawn toward selecting Black candidates, regardless of the specific qualifications of his resume. After making their selections, participants were asked to rank attributes for their importance in their decision. We expected participants to justify their decision by altering attribute weights to claim that the criterion that favored their chosen candidate was more important.

In addition, we explored the longer-term effects of casuistry, by forcing participants to engage in a second admissions task and examining how casuistry on the first selection led them to perpetuate their biased preferences through time. While Study 1 used only White participants, Study 2 included equal numbers of White and minority participants, allowing us to explore how these groups might differ when making decisions based on race. Because minority participants
are not as bound by norms of political correctness as White participants, we expected them to feel less compelled to justify their preferences.

**Method**

*Participants.* Seventy-one University of South Florida undergraduates (16 males, 55 females) received course credit for their participation. 48% were White, 20% were Hispanic, 18% were Black, 3% were Asian, and 11% reported “other.” In analyses below, we compare White participants to all other categories combined.

*Procedure.* Participants were instructed to imagine that they were on the Admissions Board at a top-ranked university, and they had one final slot to fill. In order to make political correctness norms salient, all participants read a paragraph that subtly primed the concept that some college admissions procedures were known to be biased. They received hypothetical resumes from two male high school seniors, and they were instructed to look over the resumes and select one candidate for admission. Each resume included a photograph of the candidate, information about GPA, SAT scores, number of AP classes, letters of recommendation and essays received, and various extracurricular activities. The resumes were designed to make the candidates as equally desirable/qualified as possible: The two candidates had equal cumulative SAT scores, participated in the same number of extra-curricular activities, and had the same number of academic awards. The only substantive difference was that one candidate had a superior grade point average (4.0 vs. 3.6) whereas the other candidate had taken more AP classes (9 vs. 6). In one condition, the Black candidate had the higher GPA (and therefore fewer AP classes) while in the other the qualifications were flipped such that the Black candidate had the lower GPA (and therefore more AP classes). We selected GPA and AP classes because our
previous research showed that the two were ranked similarly in importance as qualifications for college admission (Norton et al., in press).

After ranking candidates, participants were asked to rank nine criteria (letters of recommendation, GPA, student government, number of AP classes, SAT verbal, SAT math, athletic participation, essays, and race) in order of their importance in their decision. We were most interested in the relative rankings given to GPA and number of AP classes; in addition, including race as a criterion allowed us to demonstrate that participants were unwilling to acknowledge race as a criterion of importance.

To test whether participants continued to use the same criteria in further selections, we then gave participants a surprise second decision between two additional candidates. Both of these candidates were White: One had a higher GPA (3.58) but fewer AP classes (2), while the other had a lower GPA (3.32) but more AP classes (5). We could thus test to see whether participants would be consistent in selecting the candidate whose qualifications matched that of the candidate they chose in the first selection.

Results and Discussion

First Selection. As we expected, participants showed an overwhelming preference for the Black candidate, selecting him 92% (65/71) of the time, $\chi^2 (1, N = 71) = 49.03, p < .001$, and this strong preference was true for both White (88%, 30/34) and minority (95%, 35/37) participants. Most importantly, this preference held whether the Black candidate had taken more AP classes (83%, 29/35) or had the higher GPA (100%, 36/36), and these numbers were similar for both White (73% and 100%) and minority (90% and 100%) participants. In light of the fact that pretesting showed that GPA and number of AP classes were rated similarly for importance, and
therefore choice in all cells should be at 50%, these shifts in preference for the Black candidate are quite large.

*Justification.* We next checked to see if participants altered their attribute weights to support this preference. We created a dichotomous variable by coding whether participants had ranked GPA or number of AP classes higher. There was an overall preference for GPA as being more important, as 73% (52/71) ranked it higher, \( \chi^2(1, \ N = 71) = 15.34, \ p < .001 \). Importantly, however, the importance of GPA was moderated by our manipulation: When the Black candidate had the higher GPA, some 91% (32/35) ranked it more highly, but this number dropped to 56% (20/36) when the Black candidate had more AP classes, \( \chi^2(1, \ N = 71) = 11.65, \ p < .01 \). Again, results were similar for White (93% and 47%) and minority (90 and 65%) participants.

*Use of Race as a Factor.* Despite the fact that participants showed a clear bias in favor of the Black candidate regardless of his qualifications, participants ranked race as the least important (\( M = 8.09, \ SD = 1.82 \)) of the 9 attributes, significantly lower than athletic participation, the next lowest ranked qualification (\( M = 6.57, \ SD = 1.99 \)), paired \( t(71) = -5.30, \ p < .001 \). Interestingly, White participants (\( M = 8.57, \ SD = .92 \)) ranked race lower than minority (\( M = 7.65, \ SD = 2.29 \)) participants, \( t(69) = 2.20, \ p < .04 \), suggesting that minorities were more comfortable citing race as a factor in their decision-making. While results for Whites and minorities were similar for both the first selection and justification, these results begin to suggest that the two groups may approach the task differently.

*Second Selection.* These differences between Whites and minorities were revealed when we examined choices in the second selection, between two white candidates. We simply coded whether participants selected the candidate who was dominant on the same attribute as their first choice, or on the other attribute. Overall we saw some evidence for consistency, as 58% (41/71)
selected based on the same attribute. When we broke this analysis down by race, however, we saw that while White participants were very likely to be consistent (68%, 23/34), $\chi^2 (1, N = 34) = 4.24, p < .05$, minority participants were not bound by their previous decision, as only 49% (18/37) were consistent, not different from chance, $\chi^2 (1, N = 37) = .03, ns$.

**Discussion**

We observed an overwhelming preference for the Black candidate, regardless of his qualifications, suggesting the power of social category information in biasing preferences. As expected, participants altered the weights of attributes in order to select the Black candidate, inflating the value of whichever qualification favored this candidate. Participants were unwilling to admit that race had played any role in their politically correct decisions, though the selection results clearly showed that this attribute was the most important factor. Both White and minority participants favored Black candidates and inflated the value of attributes that favored this candidate, but only White participants – more strongly bound by norms of political correctness – carried them forward to a subsequent decision.

**General Discussion**

Political correctness requires both preference for minority candidates coupled with a simultaneous proscription on citing minority status as a reason for that preference. The present investigation showed that the fear of being seen as racist through politically incorrect behavior can have a powerful impact on the formation, justification, and perpetuation of preferences. Study 1 showed that even when licensed by an authority figure to cite race as a factor in decision-making, participants were highly unlikely to do so – and those participants who were particularly concerned with appearing nonprejudiced were particularly unlikely to do so. Participants in Study 2 showed an overwhelming preference for Black candidates over equally
qualified White candidates for admission to college, but ranked racial considerations last in importance for influencing their decisions, instead creatively revaluing other attributes to justify their preferences. Finally, these altered preferences carried forward in time to subsequent decisions, at least for our White participants, the social group particularly concerned with violating norms of political correctness.

Implications for Understanding Preferences

In the introduction we suggested that these studies approach preference formation in a novel way, by exploring how people form and maintain preferences when they are motivated to hide the true source of that preference. Results from Study 2 showed that people accomplished this goal by reweighting other nonracial attributes to justify that choice. Interestingly, if we look only at the qualifications of the candidates participants chose and their post-task rankings of attributes, it would be impossible to see bias – most people who picked a candidate with a higher GPA later said that GPA was more important, for example, which might signify that people are simply choosing based on their prior preference. Only by varying which qualification was paired with which candidate were we able to observe how people justified their bias. Much research has been devoted to showing that the degree to which people’s stated attribute weights actually predict their choices is much lower than would be expected (see Slovic & Lichtenstein, 1971). In an interesting experiment, Barlas (2003) showed that one reason for this lack of correspondence may be due to people’s unwillingness to give high weights to “irrational” attributes – in this case, choosing contraceptives based on pleasure or convenience rather than more rational criteria such as risk prevention. In some sense, participants in Study 2 are behaving in the same way, but while people may be reluctant to cite pleasure as a reason for their choice of contraceptives, people are actually proscribed from citing race, due to the repercussions for politically incorrect
behavior. Taken together, the present investigation and Barlas (2003) suggest that social concerns about justifying preferences can have large impacts on the formation and perpetuation of preferences, as individuals attempt to justify their choices in socially acceptable ways.

Reducing Bias

It is certainly possible that the effects we observe in Study 2 – selecting the Black candidate, then altering attributes to justify that choice, then continuing to select based on those altered attributes – are a fully conscious effort by participants to be politically correct. We would suggest that casuistry is not simply a strategy for public consumption, but that desires to mask questionable decision-making truly alter people’s preferences. We have shown that people inflate the value of attributes even prior to choice (Norton et al., in press), implying that people do not simply strategically justify their choices post-hoc, but may actually come to see attributes differently due to the power of social category information.

These private aspects of casuistry make it particularly difficult to combat: It is a challenge to debias decision-making if people are unaware they’re exhibiting bias. Study 2 revealed one means by which preferences influenced by social category can be debiased: When decision-makers were members of minority groups, and thus not as bound by norms of political correctness, the reweighting of attributes caused by social category information on the first selection did not bias subsequent choice. How might the biased preferences that our White participants formed and perpetuated be debiased? One possibility might be to have participants commit to their attribute weights prior to making their initial selection. Carlson and Pearo (2004) showed that such prior valuations of attributes (at least for consumer goods such as backpacks and wine) limited subsequent distortion of preferences. Using a similar college admissions task, we showed that such a priori weighting did not alter selection biases in favor of Black
candidates: Participants in that study simply shifted their attribute weights post-hoc to be consistent with their selection of the Black candidate (Norton et al., in press). Forced to choose between being consistent with their prior attribute rankings and following norms of political correctness and selecting the Black candidate, participants overwhelmingly chose political correctness. It is certainly possible to apply more rigorous standards for such a priori weighting of qualifications (e.g., by committing participants to an actuarial formula where ranking GPA over number of AP classes meant that GPA would account for 60% of the decision, thus committing them to choose candidates with higher GPAs). Indeed, these formulas are often used in making admission decisions; in general, however, people have strong reactions against reducing judgments about humans to regression equations (see Dawes, Faust, & Meehl, 1989), and most prefer to maintain some subjectivity in such judgments. In the domain of college admissions, for example, those who oppose actuarial methods might claim that there is no way to quantify artistic or musical ability (at least to the satisfaction of artists and musicians). Retaining subjective assessment, of course, leaves open the door for bias based on social category to tinge admissions decisions, in both directions: Those admissions officers who support affirmative action may see Black candidates as more qualified than they truly are, biasing the selection process against White applicants, while admissions officer who harbor racist beliefs may exhibit the opposite bias. While it is possible that the two opposing impulses may balance out, this seems far from an ideal solution.

*Limits of Political Correctness*

Indeed, we hasten to point out that these two opposing impulses are quite unlikely to balance out across situations in the real world. While these studies suggest that norms of political correctness can lead people to favor minority applicants for admission to college, two important
qualifiers are in order. First, our participant population (primarily White undergraduates) are among the most politically correct constituency in America, and thus may be particularly concerned about appearing nonprejudiced. Second, and related to the first point, norms of political correctness are far from the norm in America. Racism and sexism are still prevalent, of course, and countless studies have demonstrated the negative consequences of racism for minorities (see Fiske, 1998 for a review). In one recent representative study, for example, researchers sent nearly 5,000 resumes in response to help wanted ads in Boston and Chicago, and found that resumes with White names were 50% more likely to get callbacks than resumes with Black names (Bertrand & Mullainathan, 2003). Empirical studies that demonstrate favoritism toward minorities, on the other hand, are comparatively few and far between (see Dutton, 1971; Hodson et al., 2002; Norton et al., in press, for rare exceptions). While biases in favor and biases against minorities clearly do not balance out, what is most interesting about the present studies is that despite such different motivations (favoring ingroups or favoring outgroups), individuals can engage in the same strategies to justify their bias.
References


Hopwood v. Texas, 78 F.3d 932 (5th Cir. 1996).


Author Note

Michael I. Norton, Sloan School of Management, Massachusetts Institute of Technology; Joseph A. Vandello and Andrew Biga, University of South Florida; John M. Darley, Princeton University.

The authors thank Joe Simmons for his assistance with data collection, and Sam Sommers for his comments on an earlier version of this manuscript.

Correspondence concerning this article may be addressed to Michael I. Norton, Massachusetts Institute of Technology, 20 Ames St., E15-319, Cambridge, MA 02139. Electronic mail may be sent to minorton@media.mit.edu.