Marquette University

e-Publications@Marquette

Psychology Faculty Research and Publications

Psychology, Department of

2019

Being Dishonest About Our Prejudices: Moral Dissonance and Self-Justification

Kris Vasquez
Alverno College

Debra Oswald
Marquette University, debra.oswald@marquette.edu

Angela Hammer
Alverno College

Follow this and additional works at: https://epublications.marquette.edu/psych_fac

Part of the [Psychology Commons](#)

Recommended Citation

Vasquez, Kris; Oswald, Debra; and Hammer, Angela, "Being Dishonest About Our Prejudices: Moral Dissonance and Self-Justification" (2019). *Psychology Faculty Research and Publications*. 468.
https://epublications.marquette.edu/psych_fac/468

Marquette University

e-Publications@Marquette

Psychology Faculty Research and Publications/College of Arts and Sciences

This paper is NOT THE PUBLISHED VERSION; but the author's final, peer-reviewed manuscript. The published version may be accessed by following the link in the citation below.

Ethics and Behavior, Vol. 29, No. 5 (2019): 382-404. [DOI](#). This article is © Routledge Taylor & Francis Group and permission has been granted for this version to appear in [e-Publications@Marquette](#). Routledge Taylor & Francis Group does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Routledge Taylor & Francis Group.

Being Dishonest About Our Prejudices: Moral Dissonance and Self-Justification

Kris Vasquez

Department of Psychology, Alverno College

Debra L. Oswald

Department of Psychology, Marquette University

Angela Hammer

Department of Psychology, Alverno College

Abstract

We applied the moral dissonance reduction framework, used to explain the maintenance of a positive self-concept in dishonest behavior, to understand self-justification of prejudice. Participants identified ambiguously negative intergroup behaviors, then evaluated those behaviors when performed by others and themselves. As predicted by moral dissonance reduction, participants were less critical of their own behavior when considering others' behaviors before their own. In a third study directly comparing prejudiced and dishonest behavior, participants' responses showed the greatest self-

justification in the initial question about their behavior regardless of the content of the question, whereas subsequent questions showed more stability, consistent with the idea that participants adjusted their initial self-reports to avoid damage to their self-concepts.

Keywords

prejudice; self-justification; self-concept maintenance; moral dissonance

Consider the case of Valerie Smith, a candidate for village board in Southampton, New York. A reporter contacted Ms. Smith, who is White, regarding allegations that she used a racial slur in a complaint to the police about a group of Black men standing near her home. According to the reporter,

Ms. Smith, 53, confirmed that she had made the call, and repeatedly used the "n-word" throughout a phone call with a reporter. On Friday, Ms. Smith said her use of the racial slur was no big deal, and she should be entitled to use it whenever she wants. She suggested that comedian Eddie Murphy made a career out of using the word, and added that she grew up hearing it on television in the 1970s show "All in the Family." "Now, all of a sudden, I can't use it?" Ms. Smith said in a phone interview on Friday. "Sorry—I live in a black neighborhood. I came here and didn't see color. ... When you are a pioneer, like I am, it's not easy. I'm the only white person who owns and lives on this street." (Wehner, [38])

How can we understand a person who says that using a racial slur is "no big deal" and who claims to be color-blind while describing her own race and the race of the people she is slurring? One possibility is that she is an overt, "old-fashioned" racist who is comfortable explicitly embracing the idea of White supremacy. Comfort with publicly expressing prejudice may be increased by information that social condemnation of prejudice has decreased, for example, through the election of a political candidate who makes derogatory statements about minority groups (Bursztyn, Egorov, & Fiorin, [4]). There has been a substantial increase in overtly prejudiced actions following the 2016 U.S. presidential election (Levin & Grisham, [21]; Southern Poverty Law Center, [35]), the most recent national election before Ms. Smith's remarks. However, Ms. Smith does not seem to be proudly proclaiming that she and other White people are superior to other races; rather, she seems puzzled about why her use of racial slurs is problematic and defensive about her motives.

Another possibility is that Ms. Smith is expressing negative racial attitudes in a subtler way. If this is the case, she and others like her know that according to societal standards, it is a bad thing to be prejudiced. As a result of either social pressure or an internalized goal to avoid prejudice, their negative beliefs morph from a direct, old-fashioned expression to something different, such as an expression of political views that just happen to disadvantage certain racial groups or an exaggerated response to behaviors by members of that group. The idea that social condemnation can drive intergroup negativity underground is well supported; conceptions of racism that try to explain negativity combined with a reluctance to express that negativity overtly include modern racism (McConahay, [24]), aversive racism (Dovidio & Gaertner, [8]), ambivalent racism (Katz & Hass, [18]), and symbolic racism (Sears & Henry, [30]). Although these formulations differ in the way that negativity "leaks out" in judgments and interactions, and the degree to which majority group members

sincerely desire to avoid prejudice, they share the perspective that people know that expressing overt prejudice will be socially condemned. These models, however, generally would not account for Ms. Smith's behavior; none of them address behavior such as using racial slurs, which is a fairly direct expression.

A related potential explanation is that people like Ms. Smith do not know they have biases, because their biases are implicit. Implicit biases are negative evaluations of or behavioral tendencies toward outgroup members that by definition are inaccessible to introspection (Greenwald & Krieger, [17]; Greenwald & Banaji, [16]). This model shares the idea that negativity toward outgroup measures may be completely unintentional; it differs slightly in the role of individual motivation to avoid expressions of prejudice and bases in any specific moral or political stances. Instead, an individual learns passively from the environment and picks up unconscious attitudes and stereotypes. Negativity from implicit beliefs leaks into behavior in subtle ways (Pearson, Dovidio, & Gaertner, [25]) such as nonverbal behaviors and response latency (Dovidio, Kawakami, & Gaertner, [9]). However, the kinds of statements made by Ms. Smith here are quite blatant, and although it is entirely possible that she holds implicit biases and that these biases help shape her behavior and decisions, what she is saying here has a definitionally explicit component. Unlike preconscious evaluations that might play a role in decision making under uncertainty and time pressure, Ms. Smith seems to have given this matter some considerable thought. In the same interview, she brags about the effectiveness of racial slurs in disbanding crowds of her neighbors ("I said, 'You f— n—s!' and they just dispersed," Ms. Smith recalled"; Wehner, [38]).

Although there is strong research support for models explaining overt explicit prejudice, self-presentational suppression of prejudiced responses, and leakage of implicit attitudes into behavior, Ms. Smith's behavior may best be understood using a different theoretical lens: She may be defining prejudice in a flexible and self-serving way. That is, in some cases people may define prejudice as something other than what they do, no matter what it is that they do, and no matter how many people would define what they do as prejudiced. Thus a person who refuses to hire members of a minority group can argue that he is not prejudiced because he doesn't use racial slurs, where racial slurs are the mark of real prejudice; a person who does use racial slurs instead defines prejudice as something else, perhaps refusing to hire someone based on race. If people successfully redefine prejudice in this way, they may be surprised and offended that anyone would accuse them of prejudice and feel secure in their position that anyone making such accusations is wrong.[1]

A useful framework for thinking about this kind of flexible definition of wrongdoing is Ariely and colleagues' work on dishonesty (e.g., Ariely, [2]; Mazar, Amir, & Ariely, [23]). In a substantial series of studies with rich behavioral measures, these researchers have reliably induced people to lie and cheat for money. They note that very few people cheat for maximum gain even when the threat of detection is low, but many people cheat just a little bit. Researchers interpret this finding to mean that there is a range of behaviors within which people can adjust their definitions of the behaviors and maintain a positive self-concept as moral individuals while there are behaviors that are too extreme to allow for this kind of redefinition. As a result, the "small cheaters" still think of themselves as honest and ethical people, even as the overall cost of many small cheaters is much higher than that of a few serious cheaters.

In the dishonesty literature, moral dissonance reduction occurs through flexible redefinition of unethical behavior in a way that allows the maintenance of a positive self-concept (Barkan, Ayal, Gino, & Ariely, [3]). By using a self-justifying process to reduce the discomfort of moral dissonance, people can develop a "blind spot" in that they are unable to recognize their own unethical behavior. As with cognitive dissonance more generally, the individuals distort beliefs about their behavior to minimize contradictions between those beliefs and their actions; moral dissonance reduction is specifically about reducing contradictions between the unethical action and the moral code (Lowell, [22]; Shu, Gino, & Bazerman, [33]). Because the behavior is already in the public record, it is not possible to reduce the discrepancy by behaving in a way consistent with moral values, so instead, the individuals must change the way they cognitively frame the behavior, for example, by redefining it as acceptable or by adding bolstering cognitions to reaffirm their belief in their own moral virtue.

From the general literature on cognitive dissonance, we know that the self-justifying processes motivated by the discomfort of facing inconsistencies in one's thoughts and behavior (Elliot & Devine, [10]) occur so quickly as to be unnoticed by the individual. The tendency to engage in dissonance reduction gets worse when public judgment is invoked and will often lead people to pile on increasingly unlikely explanations and justifications for the original behavior (Festinger, [11]). One of the characteristic behaviors of someone engaging in dissonance reduction is their willingness to continue to defend the seemingly indefensible, using increasingly implausible arguments rather than admit to inconsistency between their actions and values. In moral and ethical domains, a person engaging in this kind of self-justification has lost the ability to see her own behavior as problematic; although her overall moral codes are intact, she has developed an ethical blind spot that prevents her from detecting her own violation of her standards (Sezer, Gino, & Bazerman, [31]).

As moral dissonance reduction occurs, the definition of dishonest behavior becomes redefined in a self-serving way that allows people to feel morally virtuous; though they are being technically dishonest, they still feel honest. Indeed, the central feature of dishonest behavior to a person in the process of dissonance reduction may well be "something that I wouldn't do." But the ability to self-justify, to reduce dissonance, is not infinite. Researchers find that the more salient the moral standards are, the less likely people are to engage in this flexible redefinition (Mazar et al., [23]); it appears that reconnecting people to their moral code makes it harder to justify even minor amounts of cheating.

Evaluations of immoral acts depend on the social context. Moral dissonance reduction is promoted by the idea that the behavior is normative (Gino, Ayal, & Ariely, [13]); thinking about how often other people do something increases the likelihood that we judge that action acceptable. At the same time, people who recall a dishonest action they have committed without the opportunity to reduce dissonance are harsher in condemning the action (Barkan et al., [3]).

We find the similarities between dishonesty and prejudice to be striking. Both behaviors are socially condemned, at least in the abstract. Both are related to moral codes. Research on prejudice reduction (e.g., Devine, Forscher, Austin, & Cox, [6]; Devine, Monteith, Zuwerink, & Elliot, [7]) argues that internalizing a nonprejudiced standard as a central moral value is key to overcoming prejudiced responses. Judgments about both categories of behavior—the degree of "wrongness" and the likelihood of behavior being excusable—are related to social norms. Social norms play a significant role in the expression or inhibition of prejudiced behaviors, particularly among those people for whom

external social pressures are significant considerations. Such people retain their underlying negative beliefs about minority groups and are prone to experiencing backlash, or resurgent expressions of negative attitudes, when the social pressure is removed (Plant & Devine, [27]).

Hypothetically, a wide variety of activities could be considered prejudiced or dishonest, and there is some debate about which behaviors don't deserve those labels; people may commit mild versions of each kind of behavior in daily living. If lying about a die roll or inflating a test score to get more money from an experimenter can be redefined as not-really-dishonest, what would a prejudice blind spot look like? By analogy, it would be an action that meets a person's abstract definition of prejudice (judging someone based on his group, treating someone badly because of his group), and it would be followed by engaging in self-justification to reduce dissonance so that the action does not feel like a moral violation to the individual. To be consistent with the dishonesty literature, we would be looking for relatively mild behaviors, not maximally prejudiced actions, as this framework implies that there are limits to what one can redefine. In a prejudice-relevant context, if moral dissonance reduction occurs, we would expect to see it in offensive intergroup behavior for which other explanations are at least somewhat plausible. For instance, after making an unfortunate comment that results in charges of prejudice, a person may engage in dissonance reduction so quickly that he is certain the comment is unobjectionable to all clear-thinking people (because he is a good person, because others have made the comment before, because the comment does not reflect who he really is, etc.).

Dishonesty researchers have argued that when we fail to understand the true nature of dishonesty—incorrectly assuming that a few "bad apples" are doing most of the harm—we cannot effectively combat the problem (Mazar et al., [23]). If moral dissonance reduction does operate in prejudice-relevant contexts, it may lead to misunderstandings about prejudiced behavior and produce similar problems. Self-justification would prevent efforts to reduce prejudice, as people with problematic attitudes and behaviors would be confident that they do not need to change. Interventions to improve intergroup relations would founder as people redefine their own actions and bolster their confidence in their self-image as a person without prejudice. But as these distortions occur, the damage inflicted by prejudiced actions would continue unabated.

The purpose of this set of studies is to investigate the ways in which prejudice can be understood from the standpoint of moral dissonance reduction. We propose that in some settings, actions that reflect negativity toward members of minority groups can be dismissed by the actor as not problematic, or at least evaluated in a way that leaves the actor feeling confident that she is not prejudiced. In the first study, we evaluate intergroup behaviors in terms of the degree of wrongness and amount of prejudice reflected for our participant population. In the second study, we use behaviors that have some degree of ambiguity as to whether they are prejudiced and provide opportunities for self-justification. In the third study, we examine the time course of dissonance reduction and make direct comparisons to dishonest behavior.

STUDY 1: EVALUATION OF NEGATIVE INTERGROUP BEHAVIORS

To assess the way prejudice fits or does not fit the moral dissonance reduction framework, we are particularly looking for potential actions that have some ambiguity to them, that may or may not be considered prejudiced. Ambiguous behaviors are much more amenable to self-justification, and in

dishonesty research these kinds of behaviors are prone to ethical blind spots (Pittarello, Leib, Gordon-Hecker, & Shalvi, [26]). Ideally, the ambiguous behaviors we identify would have the following characteristics: (a) They will not be uniformly rated as prejudicial—that is, the perception of these behaviors as prejudiced varies for individuals—and (b) they will be systematically correlated with individual differences related to moral standards related to nonprejudiced behavior and fairness to allow for the possibility of moral dissonance.

Method

Participants

This study comprised 114 Marquette University undergraduate students who participated in exchange for partial course credit. The median age was 19 years. Fifty-six percent were female, 42% were male, 1% identified as nonbinary, and 1% declined to indicate gender. Ninety percent were heterosexual, with the remainder split between homosexual and bisexual. Sixty-four percent were White, 13% Asian, 10% Latinx, 6% African American, and 6% other ethnic identities.

Materials and procedure

We provided a list of 25 intergroup behaviors that ranged in intensity from participating in a single-sex club to setting fire to a mosque. We varied the targets (race, sex, sexual orientation, disability, religion) of the behavior in the statements. We asked participants to rate the behaviors on a scale of 0 (*totally acceptable*) to 100 (*totally unacceptable*). Participants were also asked to categorize the behaviors into three groups (fine, rude/awkward, or prejudiced). The order of the questions was randomized for each participant.

We collected data on two variables that have theoretical relations to prejudice: internal and external motivation to control prejudice (Plant & Devine, [27]) and moral foundations (Graham et al., [15]). We administered the scales as developed and validated but were primarily interested in three subscales: Internal Motivation to Control Prejudice and the moral foundations of Harm and Fairness.

Internal Motivation to Control Prejudice is central to motivating the prejudice reduction process, so items that we seek to use in further studies that are related to participants' nonprejudiced identities are strong candidates for behaviors that may induce moral dissonance if committed. This subscale consists of five items measured on a 9-point scale from 1 (*strongly disagree*) to 9 (*strongly agree*) and had a Cronbach alpha of .87 in the current sample.

We include the Moral Foundations scale with an eye toward the Harm and Fairness subscales as a way of verifying that ambiguously prejudiced violations include moral considerations, and therefore that the moral dissonance framework can be appropriately applied to questions of prejudice. Each subscale consists of two parts, with three items measured on a 6-point scale, from 1 (*not at all relevant*) to 6 (*extremely relevant*), and three items measured on a 5-point scale, from 1 (*strongly disagree*) to 5 (*strongly agree*). The Harm subscale had a Cronbach's alpha of .65, and the Fairness subscale had a Cronbach's alpha of .70 in this sample.

All surveys were programmed into Qualtrics and completed online in a research laboratory.

Results

Average wrongness

The 25 statements ranged in mean wrongness on the possible scale of 0 to 100 points, from 23.8 (belonging to a single-sex club or organization) to 99.4 (beating up someone who looks gay). Standard deviations were large for all items (> 23 points for all items except the two items nearing the top of the Wrongness scale).

Categorization of behaviors

The majority of participants agreed on the category—*fine, rude or awkward but not prejudiced, or definitely prejudiced*—of 20 of the 25 items. These items are shown in Figure 1. The remaining five items had a plurality, but not majority, in one of the categories and are not included in subsequent analyses.

FIGURE 1 Mean wrongness ratings for intergroup behaviors by majority categorization.

When averaged for each participant, a within-subjects analysis of variance (ANOVA) revealed a significant effect for the ratings of wrongness, $F(2, 208) = 232.28$, $\eta^2 = .69$, $p < .001$. The estimated marginal means for all three categories were significantly different from one another at $p < .001$, using the Sidak adjustment for multiple comparisons. The wrongness of the *fine* items ($M = 31.9$, $SD = 22.8$) was markedly different from the *rude* ($M = 73.3$, $SD = 20.4$) and *prejudiced* ($M = 78.3$, $SD = 17.7$) items, the *fine* items rated relatively low on the Wrongness scale (averages of 24 to 36 points out of 100) and did not overlap at all with the wrongness ratings of the other two categories. The other two categories did have some overlap in terms of wrongness despite the overall category ratings being significantly different (see Figure 1). For instance, the *rude* items mocking a person's accent and saying "That's so gay," each averaged 80 of 100 points on the Wrongness scale. That was higher than being vigilant about safety around Black men (60 points) or being nervous near a Muslim on an airplane (72 points), items that were categorized as *prejudiced* by the majority of participants.

Relation to individual differences

We computed average wrongness scores for the preceding three groups of items and examined the correlations between the average wrongness ratings and individual differences related to prejudice and morality (Table 1).[2] The wrongness rating for the items in the *fine* group did not correlate with the individual differences of interest. The average wrongness of the *rude* and *prejudiced* items both correlated with the moral foundation of fairness; the more important that fairness was to moral

judgments for an individual, the more wrong they found the behavior. One distinction between the *rude* and *prejudiced* categories was seen in the relationship to the moral foundation of harm. The more an individual considered harm to be a central determinant to judgments about morality, the higher the wrongness ratings of the *prejudiced* items, but this relationship did not hold for the *rude* items. Both categories' wrongness correlated with internal motivation to control prejudice; those people who were more motivated to avoid showing prejudice for internalized reasons found the items more wrong than those who were less internally motivated.

TABLE 1 Correlations between Average Wrongness Ratings and Individual Difference Measures for Three Categories of Behavior

	Moral Foundation: Fairness	Moral Foundation: Harm	Internal Motivation to Control Prejudice
Fine	.08	.06	.13
Rude/Awkward	.25**	.12	.21*
Prejudiced	.35**	.26**	.38**

* $p < .05$. ** $p < .01$.

Discussion

Behaviors with negative intergroup components were interpreted on a continuum in terms of how wrong they were and whether they were considered to be clear manifestations of prejudice. Within the range of evaluations, however, there was a clear dividing line between those behaviors seen to be completely acceptable (*fine*) and those that were either rude-but-not-prejudiced or definitely prejudiced. Evaluations of the moral foundations of fairness and harm, two highly elaborated moral codes in Western societies (Graham et al., [15]), were related to judgments of how wrong the behavior was. That is, people who are more inclined to use fairness and harm in deciding whether an action is acceptable were also inclined to judge the candidate prejudice-relevant actions as more unacceptable.

The fact that perceived wrongness of the behavior was correlated with internal motivation to control prejudice is evidence that this relationship is an important precondition to applying the moral dissonance framework to prejudice. In Ariely's ([2]) work on dishonesty, one of the factors that constrains the redefinition of behavior from dishonest to honest is how much attention people pay to standards about honesty. Reminding people of their moral codes related to honesty reduces cheating. Here, moral standards are related to our list of intergroup behaviors and motivation to respond without prejudice is related to the way people judge the behaviors.

In the research on moral dissonance and dishonesty, people frequently cheat or lie a little, but very few do so for maximum possible gain. Researchers have interpreted this to mean that there are limits to how much dishonesty one can commit and still think of oneself as an honest person. Similarly, with prejudice, some behaviors appear to be clearly and universally considered prejudice, which would be difficult to deliberately engage in and still consider yourself nonprejudiced. As we seek to apply the dishonesty framework to prejudice, we suggest that the items in the *rude* category (e.g., saying "That's so gay," mocking someone's accent, or telling a boy he throws like a girl) are behaviors that have the potential to be interpreted as prejudiced, depending on the motivation and attitudes of the

participants. That is, these actions might be areas where prejudice-related self-justification is likely to occur.

STUDY 2: EXPLORING MORAL DISSONANCE REDUCTION IN PREJUDICED BEHAVIOR

We used two findings from dishonesty research to look for a potential prejudice blind spot indicating moral dissonance: People are more self-justifying when behavior is considered normative (Gino et al., [13]), and less self-justifying—even harsh upon themselves—when reflecting on their own behavior without an opportunity to reduce dissonance (Barkan et al., [3]). We see self-justifying evaluations as being the key to moral dissonance reduction. We therefore asked participants to consider one of our items in the *rude* category from Study 1 in one of two orders: thinking about other people committing the behavior and then themselves doing it, or thinking about themselves committing the behavior first and then others doing it. According to dishonesty research, people who consider the actions of others first should be more prone to self-justification. When thinking about others' behavior first, the behavior can be framed as more normative, allowing dissonance reduction to occur. If the same pattern holds for our ambiguously prejudiced items, it will provide preliminary evidence that similar moral dissonance reduction is taking place. We therefore expect people who consider others' behavior first to create room for self-justification that excuses their own behavior, represented by less condemnation of the prejudiced acts they themselves commit. We hypothesize that people will rate a behavior as worse (containing more prejudice, more morally offensive, and done less often by them compared to others) if they think about times they have committed the behavior before thinking about times that others have done so than when the order is reversed.

Method

Participants

Ninety-five Marquette University undergraduates completed the study in exchange for partial course credit. Seventy percent were female, and 29% were male. The median age was 18 years. Ninety-five percent were heterosexual, 2% homosexual, 2% bisexual, and 1% declined to answer. Sixty-five percent were White, 13% were Latinx, 10% were Asian, 6% were Black, and 6% identified with other ethnic categories.

Materials and procedure

We told participants that they would be asked to evaluate behavior that we had previously found to be relatively familiar to college students. We asked them to choose one of four possible prompts to evaluate based on what seemed most familiar to them. We did not specify whether the behavior needed to be something they themselves had done or only witnessed. The order of the choice of prompts was randomized for each participant. They could choose (a) telling a boy "You throw like a girl," (b) mocking someone's accent, (c) speaking loudly to someone with a non-hearing-related disability, or (d) using the phrase "That's so gay" to describe something dumb. Anyone who said she or he had never encountered any of these behaviors was routed out of the study. This resulted in the loss of one participant.

We randomly assigned people to one of two orders. Half of the participants were asked to write a brief description of their friends or family members doing the behavior and then rated how common this

behavior was for people in general, using a slider with a scale of 0 (*very unusual*) to 100 (*very common*). Then they were asked to write a brief description of a time they had done the behavior and rated how common it was for them personally to act this way using the same slider format. Based on the moral dissonance framework from research on dishonesty, particularly Gino and colleagues ([13]), the salience of others' behavior should promote self-justification in this condition.

The other 50% of participants, as a control, saw the prompts in reverse order (self, then others). Participants were instructed to imagine committing the behavior if they could not recall acting in the specified way. All participants then rated the behavior for how much it expressed prejudice, using a slider from 0 (*definitely not*) to 100 (*definitely*), and how wrong the action was, using a slider from 0 (*not at all wrong*) to 100 (*extremely wrong*). By first committing to record a report of their own violations, there was no opportunity for these participants to adjust their self-reports after considering the behavior of others. According to research by Barkan et al. ([3]), being forced to confront their own bad actions without the chance to excuse it away should have caused these participants less inclined to self-justify and more prone to harsh self-judgments.

After these assessments, participants completed the Moral Foundations scale (Graham et al., [15]) with a focus on the subscales related to harm ($\alpha = .58$) and fairness ($\alpha = .62$) because of their connection to egalitarian rationales for opposing prejudice. Participants were also asked whether they had reported on behavior that they personally had committed or that they imagined committing. Finally, they provided demographic data and were then routed to a separate extra-credit collector.

Sixty-three percent of participants chose to respond to the "That's so gay" prompt, 22% to making fun of an accent, 12% to telling a boy he throws like a girl, and 3% to speaking loudly to someone with a disability. Of all participants, 55% saw the "others first" condition and 45% saw the "self first" condition.

Results

Because the numbers of people responding to the different prompts were uneven, we ran each analysis for all participants and then again for the large subset who responded to the "That's so gay" prompt. The results were consistent regardless of the prompt to which they responded, so separate analyses of the specific prompts are not presented.

We computed a self–other gap score for frequency of the behavior, scored so that a positive score meant that other people are perceived to engage in the behavior more than the participant. We do not consider the ratings an accurate report of how often people really engage in these behaviors, because people are generally inaccurate at self-reports of behavior with a social desirability component, including behavior related to ethical codes (e.g., Randall & Fernandes, [28]). In this study the size of the gap varies by experimental condition, which would be unlikely if the reports were accurate and unaffected by the experimental manipulation. Although the actual ratings are not necessarily accurate representations of behavior, people's reports of how they compare themselves to other people in terms of this negative behavior are interesting. The larger the self–other gap, then, the more the participants are claiming that others act this negative way more than they do, the better they look in comparison to those others, and therefore the more self-justification is occurring.

This frequency gap was correlated with participant ratings of how prejudiced the action was, $r(95) = .27$, $p = .009$, and how wrong the behavior was, $r(95) = .31$, $p = .003$. Therefore, we ran a multivariate analysis of variance (MANOVA) with order of target order (self first or others first) as a between-subjects variable and self–other frequency gap, amount of prejudice, and wrongness rating as dependent variables. The overall MANOVA showed a significant effect of target order, $F(3, 91) = 3.11$, $\eta^2 = .09$, $p = .03$. Follow-up univariate ANOVAs showed that this effect was driven only by the self–other gap, $F(1, 93) = 7.40$, $\eta^2 = .07$, $p = .008$. There was no effect of target order on prejudice ratings, $F(1, 93) = 0.33$, $\eta^2 = .004$, $p = .57$, or wrongness ratings, $F(1, 93) = 0.06$, $\eta^2 = .001$, $p = .81$. [3]

When people thought about others' behavior first—the self-justification condition—the average gap was 41.8 points ($SD = 24.7$); when they thought about their own behavior first, the average gap was 26.4 points ($SD = 30.4$). An examination of the mean ratings shows that this effect was driven by changes in the estimates for the self. As seen in Figure 2, people estimated the frequency of others' behavior similarly when considering others' behavior first ($M = 59.5$), 95% confidence interval (CI) [53.7, 65.2], and when considering their own behavior first ($M = 55.0$), 95% CI [48.7, 61.3]. In contrast, their reports of their own behavior changed depending on whether they made those reports before or after thinking about others' behavior. Those who considered others' behavior before their own gave lower estimates of frequency for their own behavior ($M = 17.7$), 95% CI [10.8, 24.5], than those who started by thinking of their own behavior ($M = 28.6$), 95% CI [21.0, 36.1].

FIGURE 2 Mean frequency ratings for others' and own behavior by question order.

Relation to individual differences

As with Study 1, evaluations of these behaviors were related to moral reasoning related to harm and fairness. The self–other gap, an indication of self-justification, was correlated with the Moral Foundations/Harm scale, $r(95) = .27$, $p = .009$, as well as the Moral Foundations/Fairness scale, $r(95) = .33$, $p = .001$, although given the low reliability of the subscales in this sample, these results

should be viewed with caution.[4] The self–other gap was also correlated with how much prejudice people saw in the action, $r(95) = .27, p = .009$, and how wrong they thought the action was, $r(95) = .30, p = .003$. Thus the more a person viewed fairness and harm as key moral principles, the more prejudiced and wrong they judged the behavior, and the bigger the gap between how often they said they did it and how often others did it.

These correlations differed depending on the order of questions asked (self vs. others first; see Table 2). For participants who considered the behavior from the perspective of others first, in the self-justifying condition, the self–other gap was correlated with the Moral Foundations/Harm subscale, the amount of prejudice seen in the action, and the wrongness of the action. The correlation between the self–other gap and the Moral Foundations/Fairness scale was not significant. In contrast, for participants who considered their own behavior before that of others—who are predicted to be in a more self-critical state about their own behavior—the correlation between the self–other gap and the Moral Foundations/Fairness scale was significant, but the correlations between the gap and moral foundations/harm and wrongness of the action were nonsignificant.

TABLE 2 Correlations between the Self–Other Gap and Individual Differences by Order of Questions

	Others First (Self-Serving) ^a	Self First (Self-Critical) ^b
Moral foundations/Fairness	.18	.36*
Moral foundations/Harm	.36**	.16
Prejudice in action	.30*	.29
Wrongness of action	.47***	.15

a $df = 52$.

b $df = 43$.

c * $p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

As predicted from the dishonesty framework, social norms influenced the way people thought about a potentially prejudiced behavior. More self-justification occurred when people considered what others did first, consistent with previous research that others' behavior changes the salience of social norms (Gino et al., [13]). Participants in this study felt that others did the objectionable behavior more than they did across the board, consistent with the general self-serving bias (Kruger & Dunning, [20]). But when they thought about those others before thinking about their own behavior, they reported that they did these offensive actions less often than when they hadn't thought first about the actions of others (i.e., reflected on social norms). Because these order effects occurred based on random assignment, it is not plausible that the lower reports of own-frequency were accurate descriptions of group differences in behavior.

Explicit judgments of prejudice and wrongness were related to moral principles of fairness and harm. It might be argued that the correlations between the self–other gap and individual differences do not represent self-justification, and people who see the behavior as prejudiced also actually do it less, but

as previously noted, in that case the differences in reported frequency by order condition are hard to interpret. The patterns of relationships with the individual differences suggest that shifts in thinking about a specific prejudiced behavior, either in the direction of self-justification or in the direction of self-criticism, changes the matrix of beliefs about morality that a participant will report.

Although self-reports of frequency of negative behavior changed as social norms were made more or less salient, explicit judgments of the prejudice and wrongness of the behavior did not. It is difficult to interpret this finding in light of the dishonesty moral dissonance reduction framework, as to our knowledge participants in dishonesty studies are not asked how dishonest or wrong a behavior is. The lack of shift in explicit evaluations of prejudice and wrongness that we predicted in this study raise several possibilities: (a) moral dissonance reduction works differently for prejudice than dishonesty; (b) for both prejudice and dishonesty, people don't shift conscious evaluations of behavior when maintaining their self-concept—it happens at an unconscious level; (c) for both prejudice and dishonesty, people don't shift evaluations of behavior at all—behavioral data from dishonesty research may not reflect a change in the explicit definition of whether an action is dishonest; or (d) the self-serving estimation of frequency in this study reduced moral dissonance, and therefore participants had no need to further adjust their thinking in subsequent questions to maintain a positive self-concept. The goal of the final study is to tease apart these possibilities.

STUDY 3: COMPARING PREJUDICE AND DISHONESTY

In this study, we introduced two additional variables: a direct comparison between prejudice and dishonesty and the order in which reports of the frequency of behavior and the amount of prejudice/dishonesty in the behavior are measured. We were particularly interested in whether participants would demonstrate self-justifying evaluations at the first available opportunity after thinking about the immoral behavior, and whether this tendency would be different for prejudice than for dishonesty. In Study 2, people's estimates of how frequently they engaged in a behavior changed depending on whether they evaluated themselves or other first, but evaluations of how wrong the behavior was or how much prejudice it represented—questions asked after the frequency estimates—did not change. It is possible that this pattern reflects the content of the questions and that people respond to frequency estimates differently than other kinds of judgments. Another possibility is that the shift in ratings for frequency occurred simply because it was the first question encountered, and if dissonance could be reduced by giving a self-justifying estimate frequency, there was no need to further reduce dissonance by saying the actions were not that bad. In the current study we varied the content of the first question asked to test for this possibility and predicted that more self-justifying evaluations would occur in the first question asked, regardless of the content of the question.

Method

Participants

There were 301 Marquette University students who completed the study. The median age was 19 years. Seventy-seven percent were female, and 23% were male. Ninety-five percent were heterosexual, 2% homosexual, 2% bisexual, and 1% uncertain or preferred not to say. Seventy-three percent were White, 11% Asian, 7% Latinx, 3% Black, 3% biracial, less than 1% other, and 2% declined to answer their ethnicity.

Materials and procedure

This study, conducted via Qualtrics.com, was a three-way between-subjects design, manipulating the type of behavior (prejudice or dishonesty), the target order (self first or others first), and the question order (frequency of behavior first, or how prejudiced/dishonest the behavior was first). The last variable was intended to test whether participants' evaluations change in the first question asked, regardless of the topic of that question. We also asked how wrong the behavior was, though the order of this question remained constant as a control. As in Study 2, evaluations of behavior were made on a sliderbar from 0 to 100, where a higher score indicated more frequency, more prejudice/dishonesty, and more wrongness. We used the behavior most frequently recognized by participants in Study 3, using the phrase "That's so gay" to refer to something dumb, as the example of potentially prejudiced behavior, and the example of sharing answers on an academic assignment as the example of dishonest behavior.

We included a measure of moral self-concept taken from the dishonesty literature, although it should be noted that these authors (Mazar et al., [23]) predicted no differences in moral self-concept because of "self-concept maintenance." They measured self-concept maintenance by asking two questions within a battery of 10 that related to dishonesty: How honest are you (absolute scale = 0–100), and compared to other people, how moral are you (scale of 1–9, with 9 = *much more honest than other people*). We added a question of how fair you are in interactions with other people using the same scale as the honesty question. We also included the Moral Foundations scale (Graham et al., [15]), with our interest again focused on the Harm ($\alpha = .59$) and Fairness ($\alpha = .61$) subscales.

Results

In contrast to Study 2, in this study there was evidence of self-selection. People were more likely to stop completing the study or leave answers blank when asked about their own behavior first (31.1%) than when asked about others' behavior first (9.4%), $\chi^2 (1) = 28.53, p < .001$. Attrition did not differ for the two topics of prejudice and dishonesty, $\chi^2 (1) = 3.52, p = .06$.

For the participants who completed the study, we computed self–other gaps for *frequency* of behavior, amount of prejudice or dishonesty reflected in the behavior (*violation*), and *wrongness* of the behavior. The potential range of each gap was –100 to 100. Gaps were scored in each case so that a higher score is more self-justifying, that is,

- Self–other gap for frequency: Positive score means other people do it more than I do;
- Self–other gap for violation: Positive score means it is more dishonest or prejudiced when others do it than when I do it; and
- Self–other gap for how wrong the action is: Positive scores means it is more wrong when others do it than when I do it.

We conducted a three-way between-subjects MANOVA with target order, question order, and topic as between-subject factors and frequency, violation, and wrongness gaps as dependent measures. In support of the argument that perceptions of prejudice and dishonesty operate similarly, there was no main effect for the topic of the question, $F(3, 239) = 1.65, \eta^2 = .02, p = .18$, nor were there any

significant interactions involving topic. For these measures, prejudice and dishonesty evoked similar responses in participants.

There was a significant overall main effect for target order, $F(3, 239) = 3.26$, $\eta^2 = .04$, $p = .022$, as in Study 2. However, the main effect is not as theoretically interesting as the Question Order $\times$ Target Order interaction. For this interaction, the dissonance reduction hypothesis predicts that people will be more lenient in evaluating their own behavior when thinking of others' behavior before their own and that this effect will be biggest for the question asked first, as participants reduce dissonance at the first opportunity, though these results need to be interpreted in light of the unexpected differences in self-selection in different conditions. In the overall MANOVA there was a significant Target Order $\times$ Question Order interaction, $F(3, 239) = 3.10$, $\eta^2 = .04$, $p = .028$. Follow-up ANOVAs show that this effect was driven by only the violation ratings, $F(1, 241) = 8.20$, $\eta^2 = .03$, $p = .005$. When we asked the violation question first, the difference between the others-first gap and the self-first gap was considerably larger and in the predicted direction (13.0 vs. 0.4) compared to when this question was asked after the frequency question (6.5 vs. 10.0; Figure 3).

FIGURE 3 Self–other gap for violation ratings by question order.

We repeated these analyses for only those participants who said they were reporting on remembered (cf. imagined) behavior, and the results did not change. When participants first were asked about how much the behavior violates moral standards, their answers differed depending on the order of focusing on self or others first: When they were first asked about others, they judge the behavior as more violating when shown by others than when shown by themselves. When they were first asked about themselves, no differences in moral judgments appear between judging others and judging themselves. Seventy-two percent of people said they were reporting remembered behavior, although again this value should be interpreted in light of the initial self-selection.

Self-concept

We conducted a three-way between-subjects MANOVA with topic, question order, and target order as between-subject factors and participant self-rating of honesty, fairness, and ethical standards as dependent measures. If self-concept maintenance were *not* taking place, we would expect a topic main effect; people who reflected on their dishonest actions should self-rate lower on honesty and ethics, and people who reflected on their prejudiced actions should self-rate lower on fairness. The MANOVA revealed no significant effect for topic, $F(3, 290) = 0.44$, $\eta^2 = .005$, $p = .72$, nor were there other significant effects in this analysis, consistent with patterns in the dishonesty literature.

Moral foundations

The self–other gap for how frequent the behavior (across orders) was correlated with the Moral Foundations/Harm subscale, $r(268) = .13$, $p = .039$. As with Study 2, when the frequency question was asked first, the correlation between harm beliefs and the self–other gap for frequency was larger than in the overall data set, $r(133) = .22$, $p = .01$. When the violation question was asked first, the correlation was not significant, $r(135) = .04$, $p = .67$. Although, as in Study 2, the reliability of the subscales was lower than we would have liked, the consistency in the relationship between moral foundations and question order across studies in the similar condition but not in the divergent condition helps make sense of the Study 2 findings. Self-justifying estimates about how often people commit a behavior was associated with more concern about harming others as a moral issue, but only when the question about frequency was asked first. For participants for whom another question intervened between reflection on the behavior and their reports of frequency, the association between frequency and concern about harm was not present.

In contrast to Study 2, The self–other gap in frequency was not correlated with the Moral Foundations/Fairness subscale, $r(268) = .06$, $p = .31$. There were no significant correlations between this subscale and the self–other gap in frequency for either order condition.[5]

Discussion

The goal of this study was to evaluate rival interpretations of Study 2 results. We can, from these data, reject the premise that people do not shift evaluations of behavior, because the explicit evaluations of how much dishonesty or prejudice was reflected in the actions shifted in different conditions. We can tentatively reject the idea that prejudice and dishonesty are dissimilar in terms of moral dissonance reduction, at least as far as we had the opportunity to evaluate them here; in every case where it was possible for topic differences to emerge, none did. That leaves as the most likely candidate the hypothesis that self-justification through lower estimation of frequency in Study 2 reduced moral dissonance, and therefore participants had no need to further shift their evaluations of behavior to maintain a positive self-concept.

The attrition rate in the frequency-first condition in this study was a surprise and in retrospect may suggest that the selection of one of several prompts in the previous study created elements of a forced compliance cognitive dissonance design. That is, in Study 2, by choosing a prompt as most familiar, participants may have felt more obligated to answer questions about it than when we removed that free choice, even if the subsequent prompt was identical. It is impossible to know with certainty the reasons for the attrition, but its nonrandom distribution suggests that something about answering questions regarding the frequency of personally engaging in mild prejudice was uncomfortable enough

to cause significant numbers of people to cease their participation in the study. In the dishonesty literature (Shu, Mazar, Gino, Ariely, & Bazerman, [34]), signing a form prior to completing it decreased dishonesty in self-reporting by calling to mind ethical standards. Although the presentation of the frequency question was not the same as signing one's name to a form and pledging honest responses, if answering the question about one's own behavior frequency also called to mind moral standards, such salience could have increased dissonance and motivated people to opt out of the study as a way of reducing the salience of the experienced dissonance.

When looking at the participant evaluations of how much prejudice or dishonesty the behavior contained, we saw that answers depended on the order in which the questions were asked and that when the question was asked immediately after considering the behavior, ratings were consistent with the dissonance reduction hypothesis. That is, when this set of questions was the first thing people encountered after thinking about the behavior, on average they considered the behavior less dishonest or prejudiced if they thought about others' behavior before their own, allowing them to see the behavior as more normative. Thinking about their own behavior without that normative context made their evaluations harsher, as seen in studies of moral dissonance and dishonesty. In contrast, when people had the opportunity to answer other questions between thinking about the behavior and reporting on their own and others' commission of the behavior, there was not an order effect for the evaluation of how much prejudice or dishonesty the behavior reflected. The fact that the self-justifying evaluation of behavior disappears when other questions intervene between the prompt and the evaluation helps to explain the findings in Study 2 and is consistent with the idea that people are using their first responses to offload dissonance. If this is the case, there is no longer a need to self-justify when the questions come later in the sequence.

Dishonesty researchers attribute a lack of change in self-concept after cheating to self-concept maintenance (Mazar et al., [23]). Null results can be a challenge to interpret, but we see here that the two topics of prejudice and dishonesty did not differ in the participants' self-concepts after contemplating these actions, so at a minimum the findings here are consistent with published work self-concept maintenance and moral dissonance related to dishonesty.

In this study, the Study 2 correlation between the moral foundation of fairness and self-serving evaluations of frequency of behavior did not replicate. In Study 2 this correlation was strongest among those who reported the frequency of their own behavior first, which was the same group in this study that had the highest attrition rate. However, the connection between the moral foundation of harm and self-justifying reports of the frequency of behavior did replicate. There was a strong association between being concerned with harming others and being self-justifying in reporting how often one engaged in moderately unethical behavior.

GENERAL DISCUSSION

Overall, in this set of studies, we see patterns of evidence suggesting that the moral dissonance framework may be useful in understanding self-justification of some prejudiced behavior. The argument that self-justifying personal dishonesty is motivated by a desire to reduce moral dissonance, as opposed to a more general cognitive dissonance reduction desire or other processes, is based on the idea that dishonesty is considered by many people to be an immoral act. That claim is bolstered by

evidence that reminders of moral codes decrease the dishonesty being studied. Those moral codes may be about the specific behavior in question (honor codes when the behavior studied is cheating) or more general (recalling the Ten Commandments, which reduced cheating even among the nonreligious and those who recalled them incorrectly).

The analogous idea that prejudice can be reduced through making values salient has been around for nearly 50 years (Rokeach, [29]). Here, we looked at the association between specific moral foundations and reactions to behaviors in general and at the association between moral foundations and people's evaluations of their own acts. Correlations indicate that there is a relationship between how people think about morality, specifically related to themes of harm and fairness, and how they evaluate prejudice. In addition, the internal motivation to control the expression of prejudice, which requires making nonprejudiced values central to one's identity, is related to the evaluation of intergroup behaviors. For those who wish to believe themselves nonprejudiced, data from these studies, as well as from the larger literature, suggest that failure to live up to their own standards can be considered a moral issue.

Not surprisingly, there is a range of opinion on behaviors that have some negative intergroup components that could be perceived as prejudiced, and such ambiguity opens the door for self-concept maintenance. Although we would not expect dishonesty-style self-justification for the most extreme behaviors examined in Study 1, such as physical assault based on group membership, it is possible that microaggressions and other negative actions with plausible alternative explanations may allow for this kind of redefinition. People asked about the ambiguous behaviors show evidence of dissonance reduction in terms of their willingness to admit engaging in such behavior (Study 2) and their evaluations of its harmfulness (Study 3). They seem to reduce dissonance at the first available opportunity. Consistent with dishonesty research, we saw no explicit change in self-concept after the opportunity for dissonance reduction. As suggested by Shalvi, Gino, Barkan, and Ayal ([32]), dissonance reduction after a threat to one's moral self-concept is likely to occur through compensatory thoughts that reduce the degree of violation, such as the idea that other people act in worse ways.

Moral dissonance reduction related to prejudice poses a challenge for improving intergroup relations. The human capacity for self-serving justification is impressive, and the maintenance of a positive self-concept reduces the chances that people will view their behavior as problematic. Even when behavior causes problems for others, our ability to flexibly define our behavior as morally acceptable eliminates the motive to improve that behavior. To act intentionally to break the bad habit of prejudiced responses first requires one to recognize that one's actions are prejudicial in nature, whether intentional or not (Devine et al., [6], [7]); it is impossible to break a habit of which one is unaware. However, dissonance reduction has at its root a process that will prevent such recognition. Instead of mindfully seeking ways to change their own reactions, people engaged in dissonance reduction will, if pressed, simply layer on increasingly intense, less logical, justifications for the initial behavior. Such deceptive responses are not only designed to mislead others in order to protect their own reputations; these acts of dissonance reduction are also designed to self-deceive in order to maintain positive self-concepts (Chance, Norton, Gino, & Ariely, [5]).

Moral dissonance reduction further complicates the process of prejudice reduction by circumventing the necessary step of internalizing a nonprejudiced standard. Those who believe through self-

justification that they are blameless in any intergroup conflict are unlikely to feel it necessary to pledge to work on becoming less prejudiced. Self-concept maintenance will therefore prevent accurate self-assessment, and where conflicts arise, people are more likely to look outward for the source of the problem rather than inward. Ms. Smith, the politician described at the beginning of this article, argued that charges of prejudice were the result of shadowy forces opposing her candidacy to the village board rather than admit it was probably a bad idea to use the "n-word" while running for public office in a mixed-race neighborhood. She had not publicly given any signs of self-reflection or remorse, and indeed in her claims of being colorblind she has likely immunized herself from feeling the need to reduce her own prejudice while creating an impression in the minds of most observers that she is, in fact, biased (see Apfelbaum, Sommers, & Norton, [1]).

Those who believe that they are personally objective in their judgments of others, and not swayed by concepts like race or gender, show higher levels of discriminatory behavior (Uhlmann & Cohen, [37]). For such individuals, the potential for future change is reduced because those in the grip of dissonance reduction are not able to accurately predict what they will do the next time a similar circumstance arises (Chance et al., [5]). The blind spot created by this sort of ethical dissonance reduction in one instance has the potential to become the basis for self-justification in the future. In this cycle of self-justification, a person may become willing and able to justify increasingly biased and overt behavior.

Within groups, moral dissonance reduction may increase problematic behavior. Dishonesty researchers have found that the salience and social acceptability of unethical behavior, signaled by the actions of others, increases an individual's willingness to act unethically (Gino et al., [13]). To the degree that similar processes are operating for prejudice, any biased comment or action makes it easier for an observer to engage in the same kind of bias without compunction. This effect is particularly strong when the observer identifies with or feels psychologically close to the actor (Gino & Galinsky, [14]). Writing about moral dissonance more generally, but in a way certainly applicable to bias-related dissonance reduction, Lowell ([22]) observed that dissonance reduction "can create an amplifying feedback loop and downward spiral of immoral behavior" (p. 17).

Conversely, research on hypocrisy and dissonance suggests that making people aware of the discrepancies between their behavior and their professed views may assist in motivating behavioral change (Stone & Focella, [36]). Assuming that dissonance cannot be conveniently offloaded by some of the processes seen in dishonesty research, hypocrisy creates a threat to the self-integrity of the individual and motivates them to look for ways to avoid such threats in the future. It is even hypothesized that vicarious hypocrisy, seeing someone else's behavior that is inconsistent with professed beliefs, may serve the same motivational function to the degree that the observer shares an ingroup identity with the observed actor (Focella, Stone, Fernandez, Cooper, & Hogg, [12]). The key issue in applying these ideas to reduce prejudice is how to prevent people from reducing dissonance on their own so that the motivation to do better remains salient.

One dimension on which dishonesty and prejudice may diverge is the goal behind dissonance reduction. The goal of self-justification of dishonesty is to maximize profit while not suffering damage to the self-concept as an honest person (Mazar et al., [23]). Given the social stigma incurred by being labeled as prejudiced, it is likely that self-concept maintenance will be invoked when a person needs to defend the self-concept from accusations of prejudiced behavior, but it isn't clear from these studies

what the motive of the expression of prejudice is. What is the prejudice-relevant analogy to maximizing profit in dishonesty research? One possibility is that the expression of prejudiced views may be a way to affirm one's social identity among groups where the views are normative, even if larger society considers the views to be prejudiced (Klein, Licata, Assi, & Durala, [19]). Although the idea of moral dissonance reduction does not require a specific motivation other than reducing consequences of inconsistency with one's moral values, further examination of the motives involved in prejudice-related self-justification would be useful in discovering ways to circumvent this process.

We will not claim that the moral dissonance framework is suitable for understanding every negative intergroup action that occurs. We do argue, however, that both practically and theoretically it calls for further inquiry. The degree to which people can harm or insult others without adjusting their moral self-concept is a dangerous insulator against self-awareness, growth, and more positive intergroup relations.

Disclosure Statement

The authors state that no outside funding source was used to support this research, and there are no potential conflicts of interest in our reporting of the results.

Footnotes

- 1 As indeed was Ms. Smith. When the story quoted here ran online, Ms. Smith posted a public comment, not disputing any of the quotations, but saying to the reporter, "After our talk this is the conversation you want to impart? ... Obviously, there are powers that be in this Village that DO NOT want me in Government. Mr. Wehner you had an opportunity to have a real conversation with the public. To make a difference. To bring together. To rise above hate and to inform" (Wehner, [38]). How using racial slurs was rising above hate was not evident in her comments.
- 2 The remaining subscales, as predicted, were not significantly correlated with the wrongness ratings of the fine, rude, and prejudiced items. Correlation coefficients for external motivation to control prejudice were $-.11$, $.04$, and $-.08$, respectively. For the moral foundation of loyalty, coefficients were $-.15$, $-.18$, and $-.14$; for authority, $-.16$, $-.07$, and $-.16$; for purity, $-.08$, $-.17$, $-.14$.
- 3 We also ran this analysis just with people who were reporting on the behavior that they remembered doing, rather than those who said in the manipulation check that they were imagining that they did the behavior. The results were the same: a significant only effect for target order, $F(1, 60) = 6.40$, $p = .014$, driven by changes in reports of their own behaviors.
- 4 The self–other gap was uncorrelated with moral foundations of loyalty ($r = .04$, $p = .73$), authority ($r = .10$, $p = .40$), or purity ($r = -.004$, $p = .97$).
- 5 As expected, the remaining three Moral Foundations subscales were uncorrelated with the self–other gap for frequency. The gap was uncorrelated with loyalty ($r = .05$, $p = .40$), authority ($r = .06$, $p = .32$), or purity ($r = .03$, $p = .66$) in the overall data set, and no correlations emerged as significant when looking by order condition.
- 6 Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hebh.

References

- Apfelbaum, E. P., Sommers, S. R., & Norton, M. I. (2008). Seeing race and seeming racist? Evaluating strategic colorblindness in social interaction. *Journal of Personality and Social Psychology*, 95, 918. doi: 10.1037/a0011990
- Ariely, D. (2012). *The (Honest) truth about dishonesty: How we lie to everyone—especially ourselves*. New York, NY : HarperCollins.
- Barkan, R., Ayal, S., Gino, F., & Ariely, D. (2012). The pot calling the kettle black: Distancing response to ethical dissonance. *Journal of Experimental Psychology: General*, 141, 757 – 773. doi: 10.1037/a002758
- Bursztyn, L., Egorov, G., & Fiorin, S. (2017, May). From extreme to mainstream: How social norms unravel (NBER Working Paper No. 23415). Retrieved from <http://0-home.uchicago.edu.libus.csd.mu.edu/~bursztyn/Bursztyn%5fEgorov%5fFiorin%5fExtreme%5fMainstream%5f2017%5f05%5f07.pdf>
- Chance, Z., Norton, M. I., Gino, F., & Ariely, D. (2011). Temporal view of the costs and benefits of self-deception. *Proceedings of the National Academy of Sciences*, 108 (Supplement 3), 15655 – 15659. doi: 10.1073/pnas.1010658108
- Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, 48, 1267 – 1278. doi: 10.1016/j.jesp.2012.06.003
- Devine, P. G., Monteith, M. J., Zuwerink, J. R., & Elliot, A. J. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology*, 60, 817 – 830. doi: 10.1037/0022-3514.60.6.817
- Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. *Advances in Experimental Social Psychology*, 36, 1 – 52.
- Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, 82, 62. doi: 10.1037//0022-3514.82.1.62
- Elliot, A. J., & Devine, P. G. (1994). On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *Journal of Personality and Social Psychology*, 67, 382. doi: 10.1037/0022-3514.67.3.382
- Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2). Stanford, CA: Stanford University Press.
- Focella, E. S., Stone, J., Fernandez, N. C., Cooper, J., & Hogg, M. A. (2016). Vicarious hypocrisy: Bolstering attitudes and taking action after exposure to a hypocritical ingroup member. *Journal of Experimental Social Psychology*, 62, 89 – 102. doi: 10.1016/j.jesp.2015.09.014
- Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychological Science*, 20, 393 – 398. doi: 10.1111/j.1467-9280.2009.02306.x
- Gino, F., & Galinsky, A. D. (2012). Vicarious dishonesty: When psychological closeness creates distance from one's moral compass. *Organizational Behavior and Human Decision Processes*, 119, 15 – 26. doi: 10.1016/j.obhdp.2012.03.01
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, 101, 366 – 385. doi: 10.1037/a0021847

- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4 – 27. doi: 10.1037/0033-295X.102.1.4
- Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California Law Review*, 94, 945 – 967. doi: 10.2307/20439056
- Katz, I., & Hass, R. G. (1988). Racial ambivalence and American value conflict: Correlational and priming studies of dual cognitive structures. *Journal of Personality and Social Psychology*, 55, 893. doi: 10.1037/0022-3514.55.6.893
- Klein, O., Licata, L., Assi, A. E., & Durala, I. (2003). "How European am I?" Prejudice expression and the presentation of social identity. *Self and Identity*, 2, 251 – 264. doi: 10.1080/15298860309025
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77, 1121 – 1134. doi: 10.1037/0022-3514.77.6.1121
- Levin, B., & Grisham, K. (2016). Special status report: Hate crimes in the United States. San Bernardino, CA : Center for the Study of Hate and Extremism; California State University. Retrieved from <http://csbs.csusb.edu/sites/default/files/SPECIAL%20STATUS%20REPORT%20Final%20Draft.pdf>
- Lowell, J. (2012). Managers and moral dissonance: Self justification as a big threat to ethical management?. *Journal of Business Ethics*, 105, 17 – 25. doi: 10.1007/s10551-011-0931-9
- Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45, 633 – 644. doi: 10.1509/jmkr.45.6.633
- McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91 – 125). San Diego, CA, US : Academic Press.
- Pearson, A. R., Dovidio, J. F., & Gaertner, S. L. (2009). The nature of contemporary prejudice: Insights from aversive racism. *Social and Personality Psychology Compass*, 3, 314 – 338. doi: 10.1111/j.1751-9004.2009.00183.x
- Pittarello, A., Leib, M., Gordon-Hecker, T., & Shalvi, S. (2015). Justifications shape ethical blind spots. *Psychological Science*, 26, 794 – 804. doi: 10.1177/0956797615571018
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75, 811 – 832. doi: 10.1037/0022-3514.75.3.811
- Randall, D. M., & Fernandes, M. F. (1991). The social desirability response bias in ethics research. *Journal of Business Ethics*, 10 805 – 817. doi: 10.1007/BF00383696
- Rokeach, M. (1973). *The nature of human values*. New York, NY : The Free Press.
- Sears, D. O., & Henry, P. J. (2003). The origins of symbolic racism. *Journal of Personality and Social Psychology*, 85, 259 – 275. doi: 10.1037/0022-3514.85.2.259
- Sezer, O., Gino, F., & Bazerman, M. H. (2015). Ethical blind spots: Explaining unintentional unethical behavior. *Current Opinion in Psychology*, 6, 77 – 81. doi: 10.1016/j.copsyc.2015.03.030
- Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-serving justifications: Doing wrong and feeling moral. *Current Directions in Psychological Science*, 24 (2), 125 – 130. doi: 10.1177/0963721414553264
- Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Ethical discrepancy: Changing our attitudes to resolve moral dissonance. In D. De Cremer & A. E. Tenbrunsel (Eds.), *Behavioral business ethics: shaping an emerging field* (pp. 221 – 240). New York, NY : Taylor and Francis Group.

- Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences*, 109, 15197 – 15200. doi: 10.1073/pnas.1209746109
- Southern Poverty Law Center. (2016). Update: 1,094 bias-related incidents in the month following the election. Retrieved from <https://www.splcenter.org/hatewatch/2016/12/16/update-1094-bias-related-incidents-month-following-election>
- Stone, J., & Focella, E. (2011). Hypocrisy, dissonance and the self-regulation processes that improve health. *Self and Identity*, 10, 295 – 303. doi: 10.1080/15298868.2010.538550
- Uhlmann, E. L., & Cohen, G. L. (2007). "I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination. *Organizational Behavior and Human Decision Processes*, 104, 207 – 223. doi: 10.1016/j.obhdp.2007.07.001
- Wehner, G. (2017, May 26,29). Southampton village board candidate used racial slur in call to police, recording shows. The Southampton Press. Retrieved from <http://www.27east.com/news/article.cfm/Southampton-Village-Surrounding-Areas/520382/Southampton-Village-Board-Candidate-Used-Racial-Slur-In-Call-To-Police-Recording-Shows>