FlashReport

The strategic pursuit of moral credentials

Anna C. Merritt [a,*], Daniel A. Effron [a,1], Steven Fein [b], Kenneth K. Savitsky [b], Daniel M. Tuller [a], Benoît Monin [a,c]

[a] Department of Psychology, Stanford University, USA
[b] Department of Psychology, Williams College, USA
[c] Graduate School of Business, Stanford University, USA

A B S T R A C T

Moral credentials establish one's virtue and license one to act in morally disreputable ways with impunity (Monin & Miller, 2001). We propose that when people anticipate doing something morally dubious, they strategically attempt to earn moral credentials. Participants who expected to do something that could appear racist (decline to hire a Black job candidate in Studies 1 and 2, or take a test that might reveal implicit racial bias in Study 3) subsequently sought to establish non-racist credentials (by expressing greater racial sensitivity in Studies 1 and 2, or by exaggerating how favorably they perceived a Black job candidate in Study 3). Consistent with prior research, a follow-up study revealed that the opportunity to establish such credentials subsequently licensed participants to express more favorable attitudes towards a White versus a Black individual. We argue that strategically pursuing moral credentials allows individuals to manage attributions about their morally dubious behavior.

© 2011 Elsevier Inc. All rights reserved.

Introduction

People are often tempted to act in ways that could make them feel or appear immoral. When people experience such temptations, we propose, they sometimes take preemptive action to demonstrate their morality so as to forestall negative attributions about their future behavior. For example, a manager might strategically mention her support of same-sex marriage so as not to seem prejudiced when she later fires a gay employee. Such preventative steps earn *moral credentials* that establish one's good character (Merritt, Effron, & Monin, 2010; Miller & Effron, 2010).

Prior research has examined the consequences of having such credentials. When people have demonstrated a lack of prejudice (e.g., by endorsing Barack Obama), they feel more comfortable favoring Whites at the expense of Blacks (Effron, Cameron, & Monin, 2009; Monin & Miller, 2001). More generally, establishing one's morality can license selfish or unethical behavior (Jordan, Mullen, & Murnighan, 2011; Khan & Dhar, 2006; Mazar & Zhong, 2010; Sachdeva, Iliev, & Medin, 2009).

We depart from prior research by examining whether people will *strategically* pursue moral credentials in *anticipation* of doing something morally dubious. We argue that people preemptively establish their morality to ensure that their future behavior does not appear immoral to others or to themselves—in other words, to manage the attributional ambiguity surrounding their future behavior (Berglas & Jones, 1978; Dutton & Lennox, 1974; Norton, Vandello, & Darley, 2004; Snyder, Kleck, & Strenta, 1979).[2] People may seek moral credentials not only when their intentions are truly nefarious, but also when they worry that their intentions could *seem* nefarious. In either case, we propose that individuals who wish or expect to act in morally ambiguous ways will seek the attributional cover (Kelley, 1973) of moral credentials. Our studies focus specifically on how participants responded to the threat of appearing racist.

Study 1

Participants evaluated candidates for a hypothetical job. In a key condition, hiring the (slightly) more qualified candidate meant passing over a Black applicant—a choice that might seem to reflect racial bias. We predicted that before reporting their hiring decision, participants in this condition, relative to those in control conditions, would be more likely to label others' behavior as racist in an attempt to demonstrate racial sensitivity.

[2] Some initial evidence was found concurrently to our investigation by Bradley-Geist, King, Skorinko, Hebl, and McKenna (2010, Study 5). Participants were somewhat more likely to choose to describe a positive experience with a Hispanic friend vs. stranger when they later expected to write an essay opposing affirmative action, but this difference did not reach standard levels of significance.

Method

Participants

Sixty-one non-Black students (40 females) participated in exchange for course credit. One additional participant was excluded due to suspicion.

Procedure

Participants viewed information about two male job applicants, one of whom was slightly more qualified than the other. The race of the candidates, indicated by a photograph, varied by condition. In the White Qualified condition, the more qualified candidate was White and the inferior candidate was Black; in the Black Qualified condition, the reverse was true; and in the control condition, both candidates were White. Most participants (96.7%) in a separate pilot study favored the more qualified candidate, but we suspected that making this choice would nonetheless raise concerns about appearing racist when the slightly less qualified candidate was Black.

Participants were told that because the study concerned the effects of time delay on decision-making, they would make their hiring decision 24–72 h in the future. In the meantime, they were asked to complete a brief "background questionnaire" (the dependent measure), which a group of their peers would ostensibly view along with their subsequent hiring decision. As part of the questionnaire, participants indicated whether they thought each of five ambiguous behaviors was "racist" or "not racist" (e.g., "A police officer stops a Black male whose clothing and hair match the description of a crime suspect," Crosby & Monin, 2010)—judgments that afforded an opportunity to demonstrate racial sensitivity and thus garner non-racist credentials.

Results

We expected participants in the White Qualified condition to take greater steps to accumulate moral credentials than participants in the other two conditions. Consistent with this prediction, judgments of racism differed among conditions, $F(2, 58) = 4.85$, $p < .05$ (Table 1, left). A planned contrast revealed, as hypothesized, that participants in the White Qualified condition labeled significantly more of the behaviors "racist" than did participants in the other two conditions, $F(1, 58) = 6.30$, $p < .05$. Responses in the Black Qualified and control conditions did not differ significantly, $F(1, 58) = 1.61$, $p > .20$. Thus, participants expressed greater racial sensitivity when they anticipated making a decision that could appear prejudiced, presumably in an attempt to earn moral credentials that would forestall negative attributions about the decision.

Study 2

An alternative account for the results of Study 1 is that merely seeing a less-qualified Black candidate primed thoughts of racial inequality, which made participants perceive greater racism in our dependent measure. To address this alternative, Study 2 added a condition in which the White candidate was overwhelmingly more qualified than

Table 1
Mean number of behaviors labeled as "racist" as a function of the hiring decision participants expected to make in the future.

Condition	Study 1	Study 2
White Qualified	2.25 (1.25)	2.03 (1.17)
Control	1.70 (1.08)	–
Black Qualified	1.27 (1.03)	1.47 (1.07)
White Very Qualified	–	1.63 (1.03)

Note. Scores could range from 0 to 5. Standard deviations are in parentheses.

the Black candidate. This condition, if anything, should prime racial inequality more powerfully than the condition in which the White candidate was only slightly more qualified. Yet because it would be difficult to attribute a preference for an unambiguously superior White candidate to racism, this condition should not motivate participants to establish moral credentials—and thus should not increase the number of behaviors participants label as racist relative to the control condition.

Method

Participants

Ninety-seven non-Black students (46 females) participated.

Procedure

Participants were randomly assigned to the White Qualified or Black Qualified conditions from Study 1, or to a White Very Qualified condition in which the White candidate was much more qualified than the Black candidate (the Black candidate's qualifications were degraded for this condition; the White candidate's remained the same). Otherwise, the procedure was identical to Study 1.

Results and discussion

Women labeled significantly more behaviors as racist ($M = 2.00$, $SD = 1.05$) than did men ($M = 1.52$, $SD = 1.13$), $t(95) = 2.16$, $p < .05$; accordingly, subsequent analyses control for gender. Judgments of racism differed significantly among conditions, $F(2, 93) = 3.14$, $p < .05$ (Table 1, right). Planned orthogonal contrasts revealed that, as predicted, participants in the White Qualified condition labeled more behaviors as racist than did participants in the other two conditions, $F(1, 94) = 5.65$, $p < .05$, and that no significant difference emerged between the White Very Qualified condition and the Black Qualified condition, $F < 1$.

These results are inconsistent with the idea that participants construed the ambiguous behaviors as more racist because the hiring task had primed racial inequality. Participants labeled an equivalent number of behaviors as racist whether they had seen a more qualified Black candidate or a *much* more qualified White candidate. It was only when the White candidate was *slightly* superior that participants could worry that choosing to hire him might seem racist; accordingly, it was only in this condition that they sought moral credentials.

Follow-up study

We have argued that participants strategically expressed racial sensitivity in order to reduce their concern that rejecting a Black candidate would seem prejudiced. Does this strategy effectively liberate people to favor a White candidate without compunction? To find out, we had a separate group of participants ($N = 69$ White students) view the version of the materials in which the White candidate was only slightly more qualified. Next, some participants had a chance to establish credentials by identifying the ambiguous behaviors as racist, whereas others did not. All participants then used a 9-point bipolar scale to indicate which candidate they preferred to hire. Unsurprisingly, participants generally preferred to hire the White candidate ($M = 7.07$, $SD = 1.41$; higher numbers indicate a greater preference for the White candidate), as indicated by a mean significantly above the scale midpoint, $t(67) = 12.23$, $p < .001$. Yet, as predicted, this preference was stronger when participants had had an opportunity to establish credentials, ($M = 7.42$, $SD = 1.11$) than when they had not ($M = 6.75$, $SD = 1.58$), $t(67) = 2.03$, $p < .05$. Identifying behaviors as racist seems to have given participants the moral credentials they needed to feel comfortable expressing a stronger preference for the White candidate.

Study 3

Study 3 sought to replicate the strategic moral credentials effect with a more direct manipulation of the threat of appearing racist, and to address a limitation of Studies 1 and 2. Perhaps participants in the White Qualified conditions in our previous studies had already decided to hire the White candidate when they encountered the dependent measure, and accused others of prejudice to alleviate feelings of guilt surrounding their choice. In this view, participants did not label behaviors as racist to allay fears about appearing prejudiced in the future, but rather to compensate for having already done something prejudiced (cf. Jordan et al., 2011; Zhong, Liljenquist, & Cain, 2009). To minimize the possibility that a compensation effect could operate in Study 3, we led some participants to expect that they *might* receive negative diagnostic feedback about their racial attitudes in the future. This manipulation should motivate participants to establish their lack of racism preemptively without making them feel that they had already done something prejudiced.

Method

Participants

Twenty-two White students (13 female) participated.

Procedure

Participants read an adapted version of a *Washington Post* article (Vedantam, 2005) about the Implicit Association Test (IAT) and its use to measure racial attitudes. The article characterized the IAT as a measure of racial prejudice and noted that Whites who take the test are often shocked to discover that they harbor an implicit association between negative words and Black faces. In the High Threat condition, the article ended there. In the Low Threat condition, the article went on to describe the IAT as highly controversial among psychologists, some of whom argue that it merely measures benign awareness of stereotypes. Pretesting revealed that both articles made students expect to display a moderately negative association with Blacks on the IAT, but that those who read the High Threat article interpreted this association as more diagnostic of racial prejudice. Prior research has shown that completing an ostensibly diagnostic race IAT is threatening for Whites (Frantz, Cuddy, Burnett, Ray, & Hart, 2004).

After reading the article, participants learned that they would complete the IAT later in the study. Meanwhile, they were asked to rank eight applicants for a consulting job from most (1) to least (8) qualified, ostensibly for a different researcher's study. Among the information about the candidates (e.g., GPA, major) were photographs (matched for age and attractiveness) showing that seven of the candidates were White and one was Black. Our dependent measure was how favorably participants ranked the Black candidate. Participants completed this measure on a computer while the experimenter was out of the room; presumably, only the "other researcher" would view these responses.

Twelve additional White students (baseline group; five females) were recruited separately and ranked the candidates without having read either article.

Results and discussion

We predicted that participants in the High Threat condition, who were likely concerned that their IAT results would make them appear racist, would be more inclined to seek moral credentials than participants in the Low Threat condition. As predicted, participants in the High Threat condition ranked the Black candidate more favorably ($M = 3.00$, $SD = 1.48$) than those in the Low Threat condition ($M = 5.09$, $SD = 1.38$), $t(20) = 3.43$, $p < .01$. Baseline participants'

rankings of the Black candidate ($M = 5.08$, $SD = 1.54$) were nearly identical to the Low Threat participants' rankings, suggesting that the fear of appearing prejudiced led participants in the High Threat condition to exaggerate their preference for the Black job candidate. This behavior did not represent compensation for past misdeeds; instead, it seems to have been a strategic preparation for the possibility of later receiving negative, diagnostic information about one's racial attitudes.

Study 2 argued strongly against the possibility that the strategic credentials effect was due to the increased salience of racial *inequality*. Study 3 helps rule out a variant of this explanation: perhaps participants in Studies 1 and 2 labeled more behaviors as racist only when thoughts of racial *prejudice* were primed, and perhaps such priming occurred only when their hiring decision could appear racist. In Study 3, however, reading about how psychologists measure racial prejudice should have primed thoughts of prejudice regardless of how controversial this measure was purported to be. Moreover, it is not obvious that priming prejudice would lead to more favorable evaluations of a Black job candidate. In fact, in the follow-up to Study 2 described earlier, judging the racism of ambiguous statements – a task that should prime thoughts of prejudice – actually *diminished* participants' ratings of a Black candidate. We assert that the most parsimonious explanation for these findings is that participants labeled behaviors as racist (Studies 1 and 2) and exaggerated how positively they viewed a Black candidate (Study 3) in the hopes of establishing non-racist credentials.

General discussion

The present research suggests that people strategically demonstrate their morality when they fear that their future behavior could appear immoral. In Studies 1 and 2, participants expressed greater sensitivity to racism when they expected to engage in behavior that could seem prejudiced. A follow-up study showed that the opportunity to express such sensitivity licensed participants to favor a White candidate over a Black candidate more strongly. In Study 3, participants who expected to take a test that might reveal racial bias exaggerated the qualifications of a Black job candidate. Together, these studies suggest that individuals pursue moral credentials to manage the moral ambiguity surrounding their intended (Studies 1 and 2) or feared (Study 3) future behavior, and to provide attributional cover from recriminations. This research represents an important extension to work on moral licensing, which has shown that people are more likely to act in morally ambiguous ways when they have previously established their morality (Merritt et al., 2010; Miller & Effron, 2010). Past research shows that people will use moral credentials they have passively acquired; we demonstrate that people will actively seek credentials when they anticipate needing them.

One's actions are frequently judged in light of one's moral track record (Birnbaum, 1973; Effron & Monin, 2010). Anticipating this, people can actively shape their track record to shield future behavior from others' – or their own – reproach. Sometimes, people may seek credentials in order to act comfortably on legitimate motives, such as hiring the most qualified candidate regardless of race or gender. Other times, though, people may seek credentials to act on prejudicial motives with impunity.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.jesp.2011.12.017.

References

Berglas, S., & Jones, E. E. (1978). Drug choice as a self-handicapping strategy in response to non-contingent success. *Journal of Personality and Social Psychology, 36,* 405–417.

Birnbaum, M. H. (1973). Morality judgment: Test of an averaging model with differential weights. *Journal of Experimental Psychology, 99,* 395–399.

Bradley-Geist, J. C., King, E. B., Skorinko, J., Hebl, M. R., & McKenna, C. (2010). Moral credentialing by association: The importance of choice and relationship closeness. *Personality and Social Psychology Bulletin, 36*(11), 1564–1575.

Crosby, J.R., & Monin, B. (2010). If you say so: How the opinions of racial minorities influence discrimination judgments. Manuscript in preparation.

Dutton, D. G., & Lennox, V. L. (1974). Effect of prior "token" compliance on subsequent interracial behavior. *Journal of Personality and Social Psychology, 29,* 65–71.

Effron, D. A., Cameron, J. S., & Monin, B. (2009). Endorsing Obama licenses favoring whites. *Journal of Experimental Social Psychology, 45*(3), 590–593.

Effron, D. A., & Monin, B. (2010). Letting people off the hook: When do good deeds excuse transgressions? *Personality and Social Psychology Bulletin, 36,* 1618–1634.

Frantz, C. M., Cuddy, A. J. C., Burnett, M., Ray, H., & Hart, A. (2004). A threat in the computer: The race implicit association test as a stereotype threat experience. *Personality and Social Psychology Bulletin, 3,* 1611–1624.

Jordan, J., Mullen, E., & Murnighan, J. K. (2011). Striving for the moral self: The effects of recalling past moral actions on future moral behavior. *Personality and Social Psychology Bulletin, 37*(5), 701–713.

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist, 28,* 107–128.

Khan, U., & Dhar, R. (2006). Licensing effect in consumer choice. *Journal of Marketing Research, 43*(2), 259–266.

Mazar, N., & Zhong, C. -B. (2010). Do green products make us better people? *Psychological Science, 21,* 494–498.

Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral self-licensing: When being good frees us to be bad. *Social and Personality Psychology Compass, 4*(5), 344–357.

Miller, D. T., & Effron, D. A. (2010). In M. P. Zanna, & J. M. Olson (Eds.), *Psychological license: When it is needed and how it functions. Advances in experimental social psychology, 43.* (pp. 117–158) San Diego, CA: Academic Press/Elsevier.

Monin, B., & Miller, D. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology, 81,* 33–43.

Norton, M., Vandello, J., & Darley, J. (2004). Casuistry and social category bias. *Journal of Personality and Social Psychology, 87,* 817–831.

Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science, 20*(4), 523–528.

Snyder, M., Kleck, R., & Strenta, A. (1979). Avoidance of the handicapped: An attributional ambiguity analysis. *Journal of Personality and Social Psychology, 37,* 2297–2306.

Vedantam, S. (2005). *See no bias. The Washington Post* (pp. W12). January 23.

Zhong, C. -B., Liljenquist, K. A., & Cain, D. M. (2009). Moral self-regulation: Licensing & compensation. In D. De Cremer (Ed.), *Psychological perspectives on ethical behavior and decision making* (pp. 75–89). Charlotte, NC: Information Age.