RUNNING HEAD: Counterfactual Transgressions

Inventing Racist Roads Not Taken:
The Licensing Effect of Immoral Counterfactual Behaviors

Daniel A. Effron, Dale T. Miller, and Benoît Monin

Stanford University

Citation:

Effron, D. A., Miller, D. T., & Monin, B. (2012). Inventing racist roads not taken: The licensing effect of immoral counterfactual behaviors. *Journal of Personality and Social Psychology, 103,* 916-932. doi: 10.1037/a0030008

Author Note

Daniel A. Effron, Department of Psychology, Stanford University; Dale T. Miller, Graduate School of Business, Stanford University; Benoît Monin, Graduate School of Business and Department of Psychology, Stanford University.

Daniel A. Effron is now at Kellogg School of Management, Northwestern University.

This research partially constituted the first author's doctoral dissertation, and was funded in part by an American Psychological Association Dissertation Award and the Edmond J. Safra Center for Ethics at Harvard University. We thank Paul Brest, Elizabeth Mullen, and Lee Ross for insightful feedback on an earlier draft, Jennifer Eberhardt for providing the photographs used as stimuli, Rebecca Hetey for creating the morphed images used in Studies 5-6, and Tara Guarino and Kiradawn Nguyen for research assistance. Face images shown in Figure 3 are courtesy of the Face-Place Face Database Project (http://www.face-place.org/); copyright 2008, Michael J. Tarr; funding provided by NSF award 0339122.

Correspondence may be addressed to the first author, Dispute Resolution Research Center, Kellogg School of Management, Northwetsern University, 2001 Sheridan Rd. 371, Evanston, IL 60208-2001. Email: d-effron@kellogg.northwestern.edu

Abstract

Six experiments examined how people strategically use thoughts of foregone misdeeds to regulate their moral behavior. We tested two hypotheses: first, that people will feel licensed to act in morally dubious ways when they can point to immoral alternatives to their prior behavior, and second, that people made to feel insecure about their morality will exaggerate the extent to which such alternatives existed. Supporting the first hypothesis, when White participants could point to racist alternatives to their past actions, they felt they had obtained more evidence of their own virtue (Study 1), they expressed less racial sensitivity (Study 2), and they were more likely to express preferences about employment and allocating money that favored Whites at the expense of Blacks (Study 3). Supporting the second hypothesis, White participants whose security in their identity as a non-racist had been threatened remembered a prior task as having afforded more racist alternatives to their behavior than did those who were not threatened. This distortion of the past involved overestimating the number of Black individuals they had encountered on the prior task (Study 4), and exaggerating how stereotypically Black specific individuals had looked (Studies 5 and 6). We discuss implications for moral behavior, the motivated rewriting of one's moral history, and how the life unlived can liberate people to lead the life they want.

Keywords: moral credentials, licensing, counterfactual thinking, racial prejudice, counterfactual transgressions

Word Count: 12,931 (includes abstract and notes)

Inventing Racist Roads Not Taken:

The Licensing Effect of Immoral Counterfactual Behaviors

Preserving a positive identity has long been assumed by psychologists to be one of the strongest psychological motives, or one of "nature's eldest law[s]" (Allport, 1937 p. 170). For many people, being morally virtuous is a central part of this identity (Aquino & Reed, 2002). Yet people often find themselves in situations that challenge their ability to feel or appear moral. They may be tempted to act in morally dubious ways, or they may become aware that others are questioning their virtue. In such situations, people look to their past for evidence of their morality. If they can point to good deeds they have done, people will feel secure in their morality, which can allow them to undertake morally dubious behavior without worrying about the negative identity implications (Merritt, Effron, & Monin, 2010; Miller & Effron, 2010; Monin & Miller, 2001a).

We propose that people feel secure in their moral identities not only when they contemplate the virtuous paths they have chosen to tread, but also when they contemplate the immoral roads they have chosen not to take. For example, to feel secure that one is free from racial prejudice, it may not be necessary to actively support minority-group members; merely thinking about how one has not sought to harm them may suffice. Imagining *counterfactual transgressions* – bad deeds that one could have performed, but did not – allows people to think, "I'm not a bad person; if I were, then I would have transgressed." The motivated use of counterfactual transgressions, we suggest, can thus allow people to feel virtuous without performing virtuous behavior, and, ironically, to act in morally dubious ways. These predictions comport with research showing how imagining alternatives to reality (*counterfactual thinking*) can exert powerful influences on how people cope with self-threatening events (Roese, 1994;

White & Lehman, 2005), evaluate others' moral character (Miller, Visser, & Staub, 2005), and feel about their own past misdeeds (Mandel & Dhami, 2005; Niedenthal, Tangney, & Gavanski, 1994). Complementing this work, we examine how reflecting on foregone undesirable actions (counterfactual transgressions) can license less virtuous-seeming behavior.

If reflecting on their counterfactual transgressions helps people feel secure in their moral identity, two hypotheses follow. First, the easier it is for people to identify immoral alternatives to their past behavior, the more evidence of their morality they should feel they have, and the more willing they should be to act in less virtuous-seeming ways in the future. For example, situations that prompt reflection on the "racist road not taken" should increase people's feelings that they have demonstrated their lack of racism, and increase their subsequent willingness to express views that could seem racist.

Second, inducing people to feel less secure about their own morality should increase their contemplation of counterfactual transgressions. In their zeal to obtain evidence of their morality, people may even overestimate the extent to which it would have been possible to transgress. We propose that the experience or even anticipation of moral insecurity will lead people to invent immoral alternatives to their past behavior – a motivated rewriting of one's moral history. Thus, people whose circumstances increase their desire for a secure identity as a non-racist are expected to overestimate how many racist alternatives there were to their prior behavior, and how clearly racist it would have been to perform each of these alternative actions. Much like people preserve a positive self-view by distorting features of what they *have* done (e.g., exaggerating how vigorously they exercised or how frequently they performed other positive behaviors; Ross, McFarland, Conway, & Zanna, 1983; Ross, McFarland, & Fletcher, 1981), we

propose that they preserve a moral self-view by distorting features of what they have *not* done but could have done (e.g., exaggerating the number of opportunities they had to transgress).

Moral Credentials, Prior Good Deeds, and Foregone Bad Deeds

People's willingness to perform morally questionable behavior increases when they can point to past actions that secure their moral identities. For example, disagreeing with blatantly prejudiced statements (Monin & Miller, 2001a), endorsing a Black political candidate (i.e., Barack Obama; Effron, Cameron, & Monin, 2009), or writing about a positive experience with a minority-group member (Bradley-Geist, King, Skorinko, Hebl, & McKenna, 2010) licensed participants to express views that could seem prejudiced. Apparently, non-prejudiced actions made participants feel that they had *moral credentials*, and that they could thus act in dubious ways without feeling or appearing immoral (Monin & Miller, 2001a). Similarly, demonstrating one's virtue in other domains has been shown to license people to act less generously and less prosocially (Conway & Peetz, in press; Jordan, Mullen, & Murnighan, 2011; Khan & Dhar, 2006, Study 3; Mazar & Zhong, 2010; Sachdeva, Iliev, & Medin, 2009).

Complementing the tendency of people secure in their moral identity to engage in morally questionable behavior is the tendency of people *insecure* in their moral identity to seek evidence of their virtue (Dutton & Lake, 1973; Jordan, et al., 2011; Sachdeva, et al., 2009; Tetlock, Kristel, Elson, Green, & Lerner, 2000; Zhong, Ku, Lount, & Murnighan, 2010; for a review, see Zhong, Liljenquist, & Cain, 2009). This occurs whether identity insecurity arises in response to one's past misdeeds or in response to concerns that one's morality is currently or will later be in question. For example, when people expect to do something that could seem prejudiced, they strategically seek non-racist credentials by expressing more favorable attitudes towards minority-group members (Bradley-Geist, et al., 2010; Merritt et al., 2012).

We propose that a similar relationship exists between moral credentials, identity security, and the availability of counterfactual transgressions (bad acts that one could have committed but did not commit). First, to the extent that people are able to imagine immoral alternatives to their past behavior, they should feel that they have the credentials necessary to justify acting in morally dubious ways. Second, in the same way that individuals strategically establish moral credentials when they suspect they will need them (Bradley-Geist, et al., 2010; Merritt, et al., 2012), we propose that under identity threat, people will reconstruct their past to exaggerate the extent to which they could have acted (but did not act) immorally.

The Present Research

The present research examines how people use counterfactual transgressions to secure a moral identity and to regulate their behavior in the domain of racial prejudice. Although many contemporary North Americans are motivated to feel and appear non-racist (e.g., Devine, Monteith, Zuwerink, & Elliott, 1991; Dunton & Fazio, 1997; Pearson, Dovidio, & Gaertner, 2009; Plant & Devine, 1998), they often find themselves tempted to act in ways that could call their racial egalitarianism into question. Such temptations can arise from real prejudice (Dovidio & Gaertner, 1998; Gaertner & Dovidio, 1986; McConahay, 1986; Sears, 1988), or from more legitimate motives (e.g., the desire to fire an underperforming employee who just happens to be from a racial minority group). Before enacting behaviors that could seem prejudiced, people often feel the need to secure an unprejudiced identity (e.g., Crandall & Eshleman, 2003; Dutton & Lennox, 1974; Monin & Miller, 2001a). We examined whether people would draw on counterfactual acts of racism to achieve this identity security.

The present six studies test two hypotheses derived from our analysis. In Part I, Studies 1-3 examine whether the presence of salient racist alternatives to one's past behavior makes

people feel that they have evidence of their virtue (Study 1), and increases their willingness to express potentially prejudiced-seeming views (Studies 2 and 3). In Part II, Studies 4-6 examine whether people will invent prejudiced alternatives to their recent behavior when their identity as unprejudiced is insecure.

PART I:

THE LICENSING EFFECT OF THE RACIST ROAD NOT TAKEN

Study 1: Obtaining Evidence of One's Morality

Study 1 tested the hypothesis that the presence of a salient racist alternative to their prior behavior would make people feel that they had evidence of their virtue (i.e., moral credentials). We manipulated whether or not White participants had an opportunity to harm a hypothetical Black individual by accusing him of a crime he had not committed. We predicted that virtually no one would act upon this opportunity, but that participants who had been given this opportunity would feel more confident that their peers would perceive them positively.

Method

Participants. While eating in an undergraduate dining hall, 57 White participants (37 men, 23 women) completed this study on paper in exchange for candy. (An additional seven observations were excluded from participants who had already completed parts of the study or were not old enough to give informed consent).

Procedure. Participants read a vignette in which their wallet was stolen. The vignette made no mention of the thief's race, but noted several details that are associated with stereotypes of Blacks: The crime was said to have occurred in a low-income neighborhood near campus that is known to have a relatively high proportion of Black residents, and the thief was described as listening to rap music and wearing a hooded sweatshirt, jeans sagged below his hips, and

basketball shoes. These materials were intended to make it easy to imagine that a Black person had committed the crime.

The next page showed photographs of two male suspects who were said to have been detained by police for questioning. Information (adapted from Monin, Sawyer, & Marquez, 2008) presented below each photograph made one of the suspects seem guilty: He had a flimsier alibi, a record of petty theft, and was found with an unusual denomination of currency that the stolen wallet had contained (i.e., a $2 bill). We asked participants who was more likely to have committed the crime (forced choice), expecting most to accuse this guilty-seeming suspect. (The complete materials can be found in Effron, 2011).

The photograph depicting the guilty suspect always showed a White face, but we manipulated the photograph of the innocent suspect to show either a Black face (*innocent Black suspect* condition) or a White face (*two White suspects* control condition). (The manipulated photographs were matched on perceived age and attractiveness based on pilot data). Thus, only the *innocent Black suspect* condition provided participants with the option of making a racist-seeming accusation – an option that we expected few if any participants to choose.

For the dependent measures, participants imagined that other students at their university "knew nothing about [them] except which suspect [they] chose" and indicated how these students would rate them on the following semantic differentials, anchored at -3 and +3 (starred items were reverse-coded): cold/warm, moral/immoral*, prejudiced/egalitarian, open-minded/closed-minded*, likeable/dislikeable*, good/bad*, and nice/awful*. These items were averaged to form a reliable composite measure of positivity ($\alpha = .79$). We did not ask participants to rate themselves on these dimensions because prior research suggested that

estimated peer ratings would be a more sensitive measure of moral credentials (Monin & Miller, 2001b, described in Miller & Effron, 2010).

Results and discussion

All participants (26 in the *innocent Black suspect* condition and 27 in the *two White suspects* condition) accused the same guilty-seeming White suspect of the crime. To prevent an outlying observation on the dependent measure (> 3 *SD*s above the mean) from biasing analyses, we recoded it with the next-largest value in the data (see Tabachnick & Fidell, 2007). (Results were identical without this recoding, and also when the outlier was excluded). As predicted, participants thought that others would perceive them more positively when this accusation was an alternative to accusing a Black suspect ($M = .37$, $SD = .30$) than when it was not ($M = .23$, $SD = .25$) – a marginally significant difference by a two-tailed test, $t(51) = 1.95$, $p < .06$, $d = .54$.[1] Thus, participants apparently felt that not accusing a clearly innocent Black suspect provided evidence of their virtue.

Study 2: Expressing Less Racial Sensitivity

When people feel secure in the knowledge that they have evidence of their virtue, they should feel licensed to act in ways that could seem morally questionable. Thus, a salient racist alternative to people's prior behavior may license them to express less racial sensitivity. Study 2 tested this prediction using the same manipulation as Study 1. Additionally, Study 2 sought to rule out the possibility that merely priming people with racially charged information would be sufficient to produce the hypothesized licensing effect. In both the *innocent Black suspect* condition and a new *priming* control condition, participants learned that the police had detained a

[1] Comparable results were obtained when "prejudiced/egalitarian" was omitted from the composite measure, $t(51) = 2.06$, $p = .04$, $d = .58$ (one outlier was recoded as described in the main text).

guilty-seeming White suspect and an innocent-seeming Black suspect for questioning – but in the new condition, participants were not asked who they personally thought had committed the crime, and thus they had no opportunity to make a racist accusation. Although the *priming* condition allowed participants to reflect on the possible racism of detaining and even accusing an innocent Black suspect, this condition provided participants with no racist alternative to their own behavior, and thus was not expected to provide a license.

Method

Participants. Ninety-eight White participants from a predominately student population (15 of whom also identified themselves as also a member of another race; 61 females and 37 males; M_{age} = 20.10, SD = 3.57) affiliated with two different universities completed the study. (In this and in all subsequent studies, results remained significant when multiracial racial participants were excluded from analyses). Participants at one university completed the study on computers in the lab as part of a longer series of unrelated studies in exchange for $20; participants at the other university completed it online as part of a shorter series of unrelated studies in exchange for a chance to win a $50 gift card. University affiliation did not moderate the results.

Procedure. Participants read the description of the crime from Study 1, viewed the suspects detained by the police, read the relevant evidence that pointed to a particular White suspect, and indicated who had committed the crime when the innocent suspect was either White (*two White suspects* control condition) or Black (*innocent Black suspect* condition). A *priming* control condition was identical to the *innocent Black suspect* condition except that, before viewing the suspects, participants learned that they would be asked to make a judgment that was unrelated to who had committed the crime; then, after examining the suspects and the evidence,

participants were asked to judge who was "least likely to be knowledgeable about the restaurant industry." We expected that participants in the *priming* condition would accuse the same (guilty) White suspect as participants in the other conditions, because in all three conditions the other (innocent) suspect had the alibi of working at a local restaurant.

Dependent measure. After completing the manipulation, participants were directed to a different website to begin an ostensibly separate study, which actually contained the dependent measure. Salient differences in formatting further disguised the connection between the manipulation and the measure. Participants evaluated whether or not each of five ambiguous behaviors was "racist" or "not racist" (forced choice; e.g., "A woman is walking alone at night and sees a Black man coming towards her. She crosses the street;" Crosby & Monin, 2012; see also Merritt, et al., 2012).[2]

We expected that participants would privately believe that few of these behaviors represented racism. Because this opinion could seem racially insensitive, however, we expected participants to be somewhat reluctant to express it – unless they had previously had an opportunity to obtain evidence of their morality. Thus, we hypothesized that participants would label fewer behaviors as racist in the *innocent Black suspect* condition compared to the two control conditions.

Individual difference measure. Because past research found evidence that the effects of moral credentials may be strongest among relatively prejudiced individuals (Effron, 2011, Study 5; Effron, et al., 2009 Study 3), at the end of the study, participants completed a measure of racial attitudes: the six-item Modern Racism Scale (example item: "Blacks are getting too

[2] One of the original Crosby and Monin items that described the interrogation of a Black crime suspect was replaced with: "A professional comedian tells a joke that plays on negative stereotypes about Blacks."

demanding in their push for equal rights"; response options range from strong disagreement [-2] to strong agreement [+2]; McConahay, Hardee, & Batts, 1981).

Results and discussion

Exclusions. We excluded two participants from the *priming* condition: one who selected the Black suspect, and one who took greater than 4.5 *SD*s above the mean time to complete the study. This left 32 participants in the *priming* condition, 30 in the *innocent Black suspect* condition, and 34 in the *two White suspects* control condition, all of whom selected the same White suspect in response to the manipulation. Responses on the dependent measure for all remaining participants were within 3 *SD*s of the mean.

Racial sensitivity. Our dependent measure was the number of behaviors participants identified as racist. As predicted, this measure differed among the three conditions, $F(2, 93) = 3.53$, $p = .03$, $\eta^2 = .07$ (see Figure 1). A planned contrast confirmed the hypothesis that participants who had a racist alternative to their prior behavior would label fewer behaviors as racist ($M = 1.40$, $SD = .86$; coded as -2) than participants who did not (*two White suspects* control: $M = 2.00$, $SD = 1.13$; *priming* control: $M = 1.94$, $SD = .91$; each coded as +1), $F(1, 93) = 6.96$, $p < .01$, $d = .55$. The orthogonal contrast confirmed that participants in the two control conditions labeled an equivalent number of behaviors as racist, $F(1, 93) = .07$, $p = .80$, $d = .05$. (Consistent with these results, pairwise comparisons confirmed that responses in the *innocent Black suspect* condition were significantly different from both the *two White suspects* control condition, $t[93] = 2.45$, $p = .02$ $d = .51$, and the *priming* control condition, $t[93] = 2.16$, $p = .03$, $d = .45$).

Individual differences. The average score on the MRS was on the unprejudiced end of the scale ($M = -1.04$, $SD = .60$; higher numbers indicate more negative attitudes towards

Blacks). MRS did not differ significantly among conditions, $F(2, 93) < 1$, $p = .85$, nor did it significantly moderate the effects of the manipulation, $p = .28$.

Conclusion. The results suggest that being able to point to a racist alternative to one's prior behavior can license one to express less racial sensitivity. Participants who had had an opportunity to accuse a clearly innocent Black man of a crime (but did not act on it) subsequently expressed less racial sensitivity than those who had had no such opportunity. Merely priming participants with racially charged information (i.e., that the police had detained a clearly innocent Black suspect) was insufficient to produce this effect.

Study 3: Favoring Whites at the Expense of Blacks

Study 3 examined whether a racist alternative to one's prior behavior could license people to express views that favored Whites at the expense of Blacks. Participants were presented with the same manipulation used in Study 1, and then were given an opportunity to favor Whites in two hypothetical decisions: whom to hire for a job and how to divide a sum of money. We predicted that when participants could derive moral credentials from a racist alternative to their prior behavior, they would be more likely to express preferences that benefitted Whites.

Method

Participants. One hundred sixty-two White individuals affiliated with one of two universities (19 of whom also identified as members of other races; 105 females, 53 males, 4 unknown; predominately students; $M_{age} = 20.40$, $SD = 1.99$) completed Study 3 along with one or more unrelated studies at computers in the lab or online. Depending on the number of unrelated studies administered, they received $20, $5, or a chance at winning a gift card.

University affiliation did not moderate the results. (An additional 8 observations were excluded from participants who had already completed the study).

Procedure. Participants were randomly assigned to the *innocent Black suspect* condition or the *two White suspects* control condition from Studies 1 and 2. After completing the manipulation, participants began an ostensibly separate study on a different website that solicited responses to three filler items (e.g., "How knowledgeable are you about the police?") and then asked participants to imagine hiring a police officer for a job in a small town where negative attitudes about racial minorities are common. Participants indicated whether they thought the job was better suited for one race or another (-3 = *Yes, much better for a black person;* 3 = *Yes, much better for a white person*; 0 = *No, I do not feel this way at all,* i.e., no racial preference; Monin & Miller, 2001a). Next, participants indicated how they would vote to divide a $100,000 government surplus between an organization that served a primarily White community and an organization that served a primarily Black community, knowing that the latter organization had previously received a "large cash gift" from another source (11-point scale ranging in $10,000 increments from *Everything to the black group* to *Everything to the white group*; midpoint labeled *Equal amounts to both groups*; Effron et al., 2009). As in prior research, we expected that participants would be tempted to favor Whites on both measures – perhaps for non-prejudiced reasons (e.g., a Black police officer might underperform in a racially hostile environment; the Black organization might need less funding in light of the cash gift). Without non-racist credentials to reduce their fear of seeming prejudiced, however, participants should be reluctant to express a preference for favoring Whites.

Finally, participants completed the MRS as a potential moderator. At the end of the study session, they reported demographic information, including political orientation (1 = *Extremely conservative*; 7 = *Extremely liberal*).

Results

We were unable to analyze responses from two participants with incomplete data. We excluded one additional participant who was not presented with the manipulation before the dependent measures due to an error in administering the study, and two who took greater than 4.5 *SD*s above the mean time to complete the study, leaving 84 participants in the *two White suspects* control condition and 73 in the *innocent Black suspect* condition, all of whom selected the same White suspect as having committed the crime. Attrition did not differ by condition, $\chi^2(1, N = 161) = .73$, $p = .39$. For each dependent measure, outliers > 3 *SD*s away from the mean response were recoded with value of the next-largest observation (hiring measure: n = 3; funding measure: n = 1; Tabachnick & Fidell, 2007).

We hypothesized that participants who could point to a racist alternative to their prior behavior would be more likely to express preferences that favored Whites at the expense of Blacks. Consistent with this hypothesis, participants expressed a marginally stronger preference for hiring Whites in the *innocent Black suspect* condition ($M = .29$, $SD = .72$) than in the control condition ($M = .11$, $SD = .56$), $t(155) = 1.77$, $p < .08$, $d = .28$. Although participants tended to allocate a slightly greater percentage of the money White organization in the *innocent Black suspect* condition ($M = 54.45\%$, $SD = 11.31$) than in the control condition ($M = 52.26\%$, $SD = 10.79$), this difference was not significant, $t(155) = 1.28$, $p = .20$, $d = .21$.

Examining the distributions of the measures, however, raised concerns that the parametric analyses described above could be inappropriate. As is apparent in Figure 2, for both

dependent measures and in both conditions, between 50% and 75% of participants chose the scale midpoint (i.e., no racial preference). Although these distributions are unsurprising from a psychological perspective (the measures were designed to make the midpoint feel like a "safe" alternative to responses that could seem racially problematic), they represent a substantial departure from normality. To address this issue, we recoded each dependent measure to indicate whether or not each participant had expressed a pro-White preference (i.e., had or had not responded above the scale midpoint), and we performed non-parametric analyses.

The results of these analyses provided further support for the hypothesis that the manipulation would increase participants' willingness to favor Whites at the expense of Blacks. Nearly twice as many participants said they preferred to hire Whites for the police force job when there had been a racist alternative to their prior behavior (*innocent Black suspect* condition: 30.14%) than when there had been no such alternative (*two White suspects* control condition: 16.67%), $\chi^2(1, N = 157) = 4.01$, $p < .05$. Similarly, the proportion of participants who said they would allocate more money to the White community organization than to the Black organization was greater in the *innocent Black suspect* condition (39.73%) than in the control condition (27.38%), although this difference was only marginally significant, $\chi^2(1, N = 157) = 2.69$, $p = .10$. We also examined whether the manipulation affected the number of times that participants favored Whites (i.e., whether they responded above the scale midpoint on neither, one, or both items). Consistent with the hypothesis, participants in the *innocent Black suspect* condition favored Whites more times, on average, than participants in the control condition (*M*s = .70 and .44, *SD*s = .64 and .55, respectively), $t(155) = 2.73$, $p = .007$, $d = .44$.

Individual Differences. On average, participants scored on the low (unprejudiced) end of the MRS ($M = -.94$, $SD = .60$, possible range = -2 to 2). MRS scores did not significantly

moderate the results of either the parametric or the non-parametric analyses, *p*s > .24 (nor did MRS differ between conditions, $p = .35$).

Participants on average self-identified as somewhat liberal ($M = 5.17$, $SD = 1.45$; possible range: 1-7). In light of research and theory suggesting that political liberals and conservatives may have different motives for maintaining a non-prejudiced image (e.g., Nail, Harton, & Decker, 2003; Pearson, et al., 2009), we also examined whether political orientation significantly moderated the results. It did not, *p*s > .09 in the parametric analyses, and *p*s > .35 in the non-parametric analyses.

Discussion

Participants who could point to a racist alternative to their past behavior seem to have felt licensed to express views privileging Whites at the expense of Blacks. Presumably, participants felt that not accusing even a clearly innocent Black suspect attested to their virtue (Study 1), thereby allowing them to favor Whites without fear of discrediting themselves. Along with Study 2, this finding supports the hypothesis that counterfactual transgressions can license morally dubious behavior.

Unlike previous research on non-racist credentials (Bradley-Geist, et al., 2010; Effron, et al., 2009; Monin & Miller), Studies 1-3 gave participants no opportunity to express positive views about Black individuals. Our results suggest that White individuals will derive such credentials from the mere fact that it would have been possible for them to act negatively towards a Black individual. In actuality, accusing the White suspect of the crime would seem to carry little diagnostic information about one's racism, given the preponderance of evidence indicating his guilt and exculpating the Black suspect. Nevertheless, participants acted as though their refusal to take a blatantly racist action (accusing the Black suspect) provided them with

bona fides as a non-racist. To secure a non-racist identity and to license oneself to express views that could seem prejudiced, it may not be necessary to rely on the clichéd claim that "some of my best friends are Black;" instead, it may be sufficient to make the much weaker claim that "none of my worst enemies are Black."

PART II: INVENTING RACIST ROADS NOT TAKEN

Study 4: Overestimating the Number of Racist Alternatives

Studies 1-3 suggest that the presence of an immoral alternative to one's prior behavior can secure one's moral identity. Reversing the independent and dependent variables, Studies 4-6 tested the hypothesis that when people wish to secure a moral identity, they will invent immoral alternatives to their prior behavior.

Whereas Studies 1-3 gave participants moral credentials by allowing them to reflect on a racist road not taken, Study 4 tests whether participants will invent racist roads not taken if they have no other opportunities to establish moral credentials. The Study 4 paradigm began by questioning participants' racial egalitarianism in an effort to motivate them to secure a non-racist identity. Later, participants completed a task in which it was possible to make several racist judgments, none of which we expected participants to make. Following a manipulation that either did or did not provide an opportunity to secure a non-racist identity by establishing moral credentials, participants were asked to remember how many times it would have been possible to make a racist judgment in the earlier task. We predicted that without moral credentials, participants would exaggerate the number of opportunities they had had (and passed up) to make racist judgments.

Method

Participants. Sixty-two White individuals (one who also self-identified as multi-racial; 41 females and 21 males; M_{age} = 36.61, *SD* = 11.02) were recruited from a nation-wide, university-maintained, predominately non-student subject pool to complete this study on the Web in exchange for a $5 gift certificate to an online retailer.

Procedure. All participants first responded to the MRS. Whereas we administered the MRS at the end of Studies 2 and 3 as a potential moderator, we placed the MRS at the beginning of Study 4 to make participants feel that their racial egalitarianism was in question, thus motivating them to secure a non-racist identity. As described below, the manipulation subsequently allowed only some participants to satisfy this motivation.

In the next task, participants read a phrase-long description of a crime (e.g., "robbing a bank," "raping a woman on a dark street"), viewed photographs and information about three male suspects, and identified who was the criminal. As in Studies 1-3, the information we provided unambiguously pointed to the guilt of a single suspect. Participants completed this task for 16 different crimes; for each crime, the guilty suspect was White. For eleven of the crimes, both innocent suspects were also White, and in five of the crimes, one of the innocent suspects was Black (see Figure 3 for an example). We expected that most participants would accuse only the guilty White suspects, which would leave them with five occasions on which they could have instead accused clearly innocent Black suspects – that is, five racist alternatives to their prior behavior.

Moral credentials manipulation. After indicating how easy the criminal identification task had been (-3 = *Extremely difficult*, 3 = *Extremely easy*), participants were asked whether they agreed or disagreed with each of four negative statements, ostensibly as part of a separate study on "evaluating statements." For the manipulation (adapted from Monin & Miller, 2001a),

we varied the subject of these statements in order to allow some participants to reject blatantly racist sentiments and thereby satisfy the motivation of securing a non-racist identity – a motivation that completing the MRS at the beginning of the study was expected to have triggered. Participants randomly assigned to the *credentials* condition evaluated statements about "most Blacks," whereas those in two control conditions evaluated statements about either "some Blacks" or "some people:"

> [*Most blacks / Some blacks / Some people*] are more likely to be criminals than [*whites / whites / other people*].
> [*Most blacks / Some blacks / Some people*] are more violent than [*whites / whites / other people*].
> [*Most blacks / Some blacks / Some people*] belong in prison.
> [*Most blacks / Some blacks / Some people*] don't appreciate the rights they have as U.S. citizens.

We expected that participants who viewed the "most Blacks" phrasing would reject most of the statements, consequently feel that they had established non-racist credentials, and thus feel less motivated to secure a non-racist identity further. By contrast, we expected that participants who viewed the "some Blacks" or "some people" phrasing would reject fewer of the statements, consequently be denied the opportunity to establish moral credentials, and thus continue looking for opportunities to prove their own lack of racism.

Measures. Participants learned that their memory for the previously viewed suspects would be tested. After responding to filler items (e.g., how many suspects had facial hair) and receiving a reminder that they had read about 16 different crimes, participants responded to the primary dependent measure: "How many times (out of 16) did you identify a white suspect as the criminal when at least one of the other suspects in the lineup was black?" Because we expected that most participants would not have accused the clearly innocent Black suspects, this item essentially asks how many times participants could have made (but did not make) a racist

judgment. Participants were also asked how many times they had identified a Black suspect as the criminal when at least one of the other suspects shown was White. Finally, participants provided demographic information, including political orientation (1 = *Very conservative*; 7 = *Very liberal*).

We expected that participants in the credentials condition would feel more secure in their non-racist identities than participants in the two control conditions. Therefore, we predicted that participants in the two control conditions, relative to those in the credentials condition, should overestimate the number of Black suspects that they *could have* accused, but not the number of Black suspects that they *had* accused.

Results

Exclusions. We excluded four individuals whose IP addresses located them in non-English-speaking countries (e.g., China) due to concerns about language comprehension and different social norms about race, one individual who took greater than 4.5 *SD*s above the mean time to complete the study, and three who accused at least one of the Black suspects of a crime (before the manipulation). These exclusions (four in the credentials condition and four in the *some Blacks* control condition) did not differ significantly by condition, χ^2 (2, N = 62) = 4.37, p = .11. The final sample comprised 17 participants in the credentials condition, 17 in the *some Blacks* condition, and 20 in the *some people* condition. Responses on the dependent measure for all remaining participants were within 3 *SD*s of the mean.

Identification of suspects. Most participants (92%) in the sample we analyzed selected all 16 White suspects we had intended to seem guilty (the other 8% accused at least one White suspect who was not intended to seem guilty). Participants rated this task as subjectively easy (M = 2.30, SD = .85).

Manipulation check. As expected, the number of statements with which participants agreed differed significantly by condition, $F(2, 52) = 40.62$, $p < .0001$, $\eta^2 = .61$. On average, participants agreed with almost none of the four negative statements about *most Blacks* ($M = .35$, $SD = .79$), significantly more of the statements about *some Blacks* ($M = 1.59$, $SD = 1.50$), and almost all of the statements about *some people* ($M = 3.60$, $SD = .94$), ps $< .01$ for all pairwise comparisons.

Remembered opportunities to accuse Black suspects. As previously mentioned, all participants included in analyses accused only White suspects. We predicted that participants who had not established credentials would try to secure a non-racist identity by exaggerating the number of times they could have accused an innocent Black suspect instead. Consistent with this prediction, participants' memories for the number of foregone opportunities to accuse Black suspects differed significantly among conditions, as indicated by a *W*-test (Welch, 1951), F_w (2, 32.57) = 4.05, $p = .03$ (see Figure 4). (A traditional ANOVA was also significant, $F[2, 51] = 3.49$, $p = .04$, $\eta^2 = .12$, but was deemed inappropriate due to heteroskedasticity; Cook-Weisberg test: $\chi^2[1, N = 53] = 5.69$, $p = .02$). Planned orthogonal contrasts tested specific hypotheses using the Welch-Šidák heteroskedastic method (see Wilcox, 2003). An initial contrast confirmed that participants who had been given a chance to establish credentials reported that there had been fewer innocent Black suspects (*most Blacks* condition, coded as +2; $M = 5.24$, $SD = 3.19$) than participants given no chance to establish credentials (*some people* and *some Blacks* control conditions, each coded -1; Ms = 9.50 and 7.47, SDs = 6.05 and 4.77, respectively), $t(47.85) = 2.75$, $p = .008$, $d = .80$. A second contrast confirmed that participants reported remembering an equivalent number of Black suspects in the two control conditions (coded as -1 and +1; *most Blacks* condition coded 0), $t(34.83) = 1.14$, $p = .26$, $d = .39$. (Pairwise comparisons revealed that

participants' responses in the *most Blacks* condition differed significantly from those in the *some people* condition, $t[29.71] = 2.74$, $p = .01$, $d = 1.01$, and that the difference between responses in the *most Blacks* and the *some Blacks* conditions was in the predicted direction but not significant, $t[27.93] = 1.61$, $p = .12$, $d = .61$).

How accurate were participants' memories? In reality, there were 5 Black suspects whom participants did not accuse. When participants could establish credentials, their reports did not deviate systematically from this correct response, $t(16) < 1$, *ns*, in a one-sample test comparing the mean number of remembered Black suspects to 5. By contrast, when participants could not establish credentials, they overestimated the number the number of Black suspects they could have accused, $t(18) = 3.33$, $p < .005$ for the *some people* control condition, and $t(16) = 2.14$, $p < .05$ for the *some Blacks* control condition.

Remembered accusations of Black suspects. Whereas exaggerating the number of *foregone* opportunities to accuse Black suspects should help secure a non-racist identity, exaggerating the number of times one *did* accuse Black suspects should not. Thus, we predicted that participants without credentials would not exaggerate the number of Black suspects whom they *had* accused. This prediction was confirmed, although the means on this measure differed marginally among conditions in an unexpected way, $F_w (2, 24.26) = 3.17$, $p < .06$. As predicted, the planned contrasts described in the prior section revealed that participants in the *most Blacks* (credentials) condition remembered accusing an equivalent number of Black suspects ($M = .82$, $SD = 1.13$; coded +2) as participants in the two control conditions (each coded -1), $t(32.29) < 1$, $p = .96$. Unexpectedly, participants in the *some people* control condition ($M = .25$, $SD = .44$; coded +1) tended to remember accusing fewer Black suspects than participants in the *some*

Blacks control condition (M = 1.35, SD = 2.60; coded -1), but this difference was only marginally significant, $t(16.80) = 1.73$, $p = .10$, $d = .84$.

Together with the results presented in the previous section, these findings confirm the predictions that non-credentialed participants (compared to credentialed participants) would reporting having *foregone* more opportunities to accuse Black suspects, and would not report having *taken* more opportunities to accuse Black suspects.

Individual differences. Compared to the student samples in Studies 2 and 3, the Web participants in the present study had more prejudiced scores on the MRS, but the mean score was still on the less-prejudiced side of the scale midpoint (M = -.40, SD = .77; possible range: -2 to 2) and represented a greater diversity of political orientations (M = 4.07, SD = 1.72; possible range: 1-7). Neither MRS nor political orientation significantly moderated the effects of the manipulation on how many times participants thought the task had allowed them to accuse Black suspects, ps > .89, or the number of times they thought they actually did accuse Blacks suspects, ps > .09.

Discussion

As long as participants had an opportunity to reject blatantly racist statements, their memories of the suspects in the criminal identification task showed no systematic bias. Even though rejecting these statements hardly represents compelling proof that one is free from all racial bias, participants nonetheless seem to have taken comfort in having these non-racist credentials to point to (Monin & Miller, 2001). But when participants lacked even this reassuring track record, they invented one by exaggerating the number of foregone racist alternatives to their prior behavior. Credentialed participants may have made a mountain of

virtue out of a molehill of rectitude, but non-credentialed participants seem to have invented the molehill.

Study 5: Exaggerating Racial Stereotypicality

Study 5 provided another test of the idea that people will try to achieve security in their identity as non-racists by inventing racist roads not taken. Instead of exaggerating the *number* of racist alternatives available (as they could do in Study 4), participants in Study 5 could exaggerate *how clearly racist* a given alternative would have been to select. Participants completed a task in which they were expected not to accuse each of several racially ambiguous suspects of crimes. We predicted that participants who could not otherwise secure a non-racist identity (relative to those who could) would remember these innocent suspects as looking more stereotypically Black, thus implying that it would have been more clearly racist to have accused these suspects. To rule out the possibility that the manipulation would make participants remember *anyone* as stereotypically Black, additional conditions induced participants to accuse the racially ambiguous suspects. The manipulation should not have the same effect in these conditions because remembering these accused suspects as stereotypically Black would not make one seem non-racist.

Method

Participants. White participants (N = 134; 85 females, 45 males, 4 unknown; M_{age} = 37.88, SD= 11.61) were recruited and compensated as in Study 4..

Procedure. Participants first completed a measure of internal and external motivation to respond without prejudice (IMS/EMS; Plant & Devine, 1998), as well as the MRS. As in Study 4, the purpose of administering these measures first was to make participants feel that their racial egalitarianism was in question, and to motivate them to secure a non-racist identity.

Administering the IMS/EMS also allowed us to test whether the effects of a moral credentials manipulation would be more pronounced for individuals with stronger motivations to respond without prejudice, although prior research found no evidence of such moderation (Monin & Miller, 2001a).

Participants then completed a new version of the criminal identification task used in Study 4. This version used eight photographs of White male faces, taken from the materials used in Study 4, and four images of racially ambiguous male faces (i.e., neither obviously Black nor White), each of which was created by morphing together photographs of one White and one Black face (Morpheus Software, 2011). For each of four crimes, participants identified the criminal from among three suspects, depicted by two White photographs and one racially ambiguous photograph.

Manipulations. The photographs and information about the suspects remained constant across conditions, but we manipulated (between subjects) which photograph was paired with information that pointed to the guilty suspect. In the *guilty White suspect* condition, participants were expected to accuse a different White suspect for each of the four crimes, and not to accuse any of the racially ambiguous suspects. Conversely, participants in the *guilty ambiguous suspect* condition were expected to accuse a different racially ambiguous suspect for each of the crimes. Only in the former condition would perceiving the ambiguous suspect as stereotypically Black imply that one could have made but did not make a racist accusation.

After indicating how easy or difficult the task had been (see Study 4) and completing other filler items (e.g., "Have you ever been asked to identify a suspect in real life?"), participants completed the second manipulation, orthogonal to the first, which was framed as a separate study about "evaluating statements." Participants were randomly assigned to either the

credentials condition or to the *some Blacks* control condition of Study 4, with slight changes in wording. Specifically, participants were asked whether they agreed or disagreed with negative statements about either all Black people or some Black people:

> [*All/Some*] blacks are criminals.
> [*All/Some*] blacks are more violent than [*whites / some whites*].
> [*All blacks*/*There are blacks who*] belong in prison.
> [*All blacks/There are blacks who*] don't appreciate the rights they have as US citizens.

We expected that participants would be more likely to reject the statements and thus feel that they had established non-racist credentials when the statements referred to "all Blacks."

To summarize, the study represented a 2 (guilty suspect: White vs. racially ambiguous) X 2 (credentials: yes vs. no) factorial design.

Measures. For each of the four crimes, participants viewed the information about the suspects again, with the photograph of the ambiguous suspect omitted. Participants were reminded whether or not they had accused this suspect, and were then asked to identify which of six photographs depicted him. These photographs were different morphed blends of the two photographs used to create the ambiguous face, and ranged from more stereotypically White-looking (coded as 1) to more stereotypically Black-looking (coded as 6), with equivalent differences in morphing between each response option (Morpheus Software, 2011).

Next, participants indicated whether the photograph they had selected was Black or White (forced choice), and used 7-point scales to rate how certain they were about his race (1 = *Not at all certain,* 7 = *Completely certain*) and the extent to which he looked like a stereotypical member of his racial group (1 = *Not at all*, 7 = *Extremely*). Finally, participants provided demographic information.

Accuracy motivation condition. To provide a baseline for the dependent measures, we sought to create a condition that would motivate participants to remember the photographs

accurately. After we had run the conditions described earlier, we showed the photographs of the suspects to a different group of participants, but omitted all references to crimes and suspects. Individual-difference measures of prejudice (i.e., MRS, IMS/EMS) were also omitted to minimize participants' concerns that their racial egalitarianism might be in question. Participants' goal was to "try to commit each face to memory by studying it carefully for at least 10 seconds" in preparation for a memory test, and to indicate which photograph looked youngest. Finally, participants evaluated statements unrelated to race (i.e., the statements about "some people" used in Study 4), and completed the dependent measures.

Results and discussion

Exclusions. We were unable to analyze data from two participants who experienced technical problems and from one participant who did not finish the study. We also excluded two participants who took greater than 4.5 *SD*s above the mean time to complete the study. (No IP addresses were located in non-English-speaking countries). Exclusions did not differ significantly among conditions, $\chi^2(4, N = 134) = 1.99, p > .70$, and left between 25 and 27 participants in each. Responses on the dependent measure for all remaining participants were within 3 *SD*s of the mean.

Manipulation checks. As expected, the proportion of participants who accused only White suspects depended on the manipulation of suspect guilt, $\chi^2(1, N = 104) = 104.00, p < .0001$. In the *guilty White suspect* condition, 100% accused all of the White suspects who were meant to seem guilty, whereas in the *guilty ambiguous suspect* condition, 92% accused all of the ambiguous suspects.

Participants also responded as expected to the credentials manipulation: Whereas only two participants agreed with any of the four negative statements about "all Blacks" (credentials

condition; $M = .04$, $SD = .23$), participants tended to agree with most of the negative statements about "some Blacks" (control condition; $M = 3.55$, $SD = .76$), $t(102) = 32.26$, $p < .0001$, $d = 6.39$.

Remembering the ambiguous suspects' race. To create the primary dependent measure, we first averaged participants' choice of photograph across each of the four crimes ($\alpha = .62$). Next, we created scales of how certain participants were that the ambiguous suspects were Black, and how stereotypically Black they looked. This required multiplying the certainty and stereotypicality measures by -1 for suspects who participants thought were White, so that higher numbers would indicate remembering the suspect as looking "Blacker." We averaged the resulting scales (which ranged from -7 to 7) across the four ambiguous suspects ($\alpha = .74$ for the certainty scale; $\alpha = .70$ for the stereotypicality scale). Finally, because the photograph scale, the certainty scale, and the stereotypicality scale were in turn highly correlated, we standardized and averaged each to form a single composite measure of "how Black" participants remembered the ambiguous suspects as looking ($\alpha = .90$). This composite served as our primary dependent measure.

Participants should be relatively more motivated to secure a non-racist identity if they had not previously established moral credentials. One way to secure a non-racist identity would be to remember the innocent suspects (whom participants had not accused) as looking more stereotypically Black. Based on this logic, we predicted that participants in the control condition (compared to the credentials condition) would remember the racially ambiguous suspects as looking "Blacker" – but only if participants had been induced not to accuse these suspects. To test this prediction, we submitted the composite measure to a 2 (guilt: ambiguous suspect vs.

White suspect) X 2 (credentials: yes vs. no) ANOVA, which revealed the predicted interaction, $F(1, 100) = 5.64$, $p = .02$, partial $\eta^2 = .05$, and no significant main effects, ps $> .11$.[3]

Tests of simple effects confirmed the specific hypotheses (see Figure 5). When participants had been induced to accuse White suspects, they remembered the ambiguous suspects as looking "Blacker" in the no-credentials condition than in the credentials condition (Ms =.51 and -.19, SDs = .93 and 1.03, respectively), $F(1, 100) = 7.97$, $p = .006$, $d = .56$. By contrast and also as predicted, when it was the ambiguous suspects that participants had accused, the credentials manipulation had no such effect (no-credentials vs. credentials: Ms = -.02 and .12, SDs = .73 and .93, respectively), $F(1, 100) = .31$, $p = .58$, $d = .11$. Decomposing the interaction the other way revealed that participants who lacked credentials remembered innocent ambiguous suspects as significantly "Blacker" than guilty ambiguous suspects, $F(1, 100) = 4.36$, $p = .04$, $d = .42$, unlike participants who had credentials, $F(1, 100) = 1.60$, $p = .21$, $d = .25$.

Comparisons with accuracy motivation condition. The results just presented do not allow us to determine whether the stereotypical "Blackness" of the innocent suspects was overestimated by non-credentialed participants (as we predicted) or underestimated by credentialed participants. To address this question, we regressed the dependent measure on two dummy codes, each of which compared one level of the credentials manipulation to the *accuracy motivation* condition when the ambiguous suspects were innocent. Results showed that when participants could not otherwise establish credentials, they remembered the innocent suspects as

[3] A mixed model that specified guilt, credentials, and their interaction as fixed effects and participant and specific suspect as random effects produced identical conclusions. The hypothesized interaction effect also emerged separately on each of the three items in the composite measure: the photograph participants chose, $F(1, 100) = 3.70$, $p < .06$, partial $\eta^2 = .04$, how certain they were that the person in this photograph was Black, $F(1, 100) = 6.02$, $p = .02$, partial $\eta^2 = .06$, and how stereotypically Black they thought he looked, $F(1, 100) = 4.45$, $p = .04$, partial $\eta^2 = .04$ (these tests use the ANOVA model described in the main text).

significantly "Blacker" than participants in the *accuracy motivation* condition did (accuracy motivation: $M = -.43$, $SD = .67$; see Figure 5), $t(75) = 3.78$, $p < .001$, $d = .87$. But when participants *could* establish credentials by disagreeing with racist statements, their memories for the innocent suspects were statistically indistinguishable from responses in the accuracy motivation condition, $t(75) = .95$, $p = .34$, $d = .22$. Thus, as predicted, participants who could not secure a non-racist identity seem to have exaggerated how stereotypically Black the innocent suspects looked.

Individual differences. The interaction between suspect guilt and credentials on memories for the ambiguous suspect's race was not significantly moderated by MRS ($M_{mrs} = -.52$, $SD = .79$; possible range: -2 to 2), IMS ($M_{ims} = 7.44$, $SD = 1.58$; possible range: 1-9), EMS ($M_{ems} = 4.44$, $SD = 2.17$; possible range: 1-9), or political orientation ($M_{politics} = 4.15$, $SD = 1.68$; possible range: 1-7), ps > .34. (Higher numbers on the IMS and EMS indicate, respectively, a stronger internal and external motivation to respond without prejudice).

Summary. Compared to participants who had been given an opportunity to secure a non-racist identity, participants given no such opportunity remembered racially ambiguous suspects whom they had not accused of a crime as looking more stereotypically Black. By altering their reconstructed memories for the faces, participants in effect invented opportunities for racism that they would not otherwise have thought existed.

Study 6: The Anticipation of Seeming Racist

Study 6 used a more direct manipulation of identity insecurity to seek convergent support for the idea that people invent counterfactual transgressions when motivated to secure a moral identity. Studies 4 and 5 sought to motivate participants in all conditions to secure a non-racist identity, and then manipulated whether or not participants could satisfy this motivation by

establishing non-racist credentials. By contrast, Study 6 manipulated whether or not participants would have a reason to feel insecure in their non-racist identities in the first place. Rather than giving some participants a chance to reject blatantly racist statements, we varied whether participants in Study 6 expected to argue in favor of negative statements about Blacks or about teenagers. We predicted that participants who expected to make such arguments about Blacks would fear appearing racist, and would thus be motivated to secure a non-racist identity preemptively. One way to achieve such security, we predicted, would be to seize an opportunity to invent racist alternatives to one's prior behavior.

Method

Participants. Seventy-six White participants (63 females, 11 males, 2 unknown; M_{age} = 42.39, SD = 12.07) were recruited from the online subject pool used in Studies 4 and 5, and were entered into a drawing for a $50 gift card for their participation.

Between 10 and 13 months before we recruited them, participants had completed the IMS/EMS, the MRS, and another measure of racial attitudes: the Attitudes Towards Blacks (ATB) scale (Brigham, 1993). The primary purpose of administering the MRS and IMS/EMS at the beginning of Studies 4 and 5 was to motivate participants to secure a non-racist identity. By contrast, the purpose of pretesting participants on these measures plus the ATB scale months before the present study was to allow us to test for moderation by these measures without affecting motivation in the study itself.

Procedure. Participants completed the version of the criminal identification task used in Study 5. The information provided always made a White suspect seem guilty. Afterwards, participants rated how easy they found this task and completed filler items, and learned that a separate study would next examine "how people think about different perspectives on

controversial social issues." Then they were shown two statements (described below) and told that later in the study, they would choose one and "list reasons why [it] might be true." Participants read that they could draw these reasons from "your personal experience, your knowledge about society, or from any other relevant source." Further, they read, "You may pick whichever statement you want, but we encourage you to pick the one that you think is more true than the other."

Manipulation. We manipulated the relevance of the two statements to race. The *insecure identity* condition presented statements with which most participants in Studies 4 and 5 had disagreed: "Most blacks are more likely to be criminals than whites" and "Most blacks are more violent than whites." In the control condition, the statements referred to non-racial social categories: "Teenagers are more likely to commit crimes than people older than age 30" and "Teenagers are more violent than people older than age 30." We predicted that only the expectation of arguing in favor of a negative statement about Blacks would raise concerns about appearing racist and would motivate participants to secure a non-racist identity (cf. Bradley-Geist, et al., 2010). Note that because participants had completed the measures of racial attitudes months before beginning this study, we expected that the control condition would provide little motivation to secure a non-racist identity (in contrast to the no-credentials conditions in Studies 4 and 5).

Comprehension checks. Participants were told that before they chose a statement, their memory for the study so far would be assessed. Comprehension-check items asked if (a) the study required reading or writing an essay about one or both of the statements shown earlier, and (b) whether the essay would argue that the statement(s) was (were) true or false.

Dependent measures. Participants next responded to the dependent measures used in Study 5, which asked them to recall the race of the racially ambiguous suspects.

Manipulation checks. Finally, participants chose which of the two statements (described earlier) to write about. Manipulation checks asked how uncomfortable participants felt about describing why their chosen statement might be true, and how concerned they were that this writing task would make them look bad (1 = *Not at all*, 7 = *Very*). Participants wrote the essay, indicated how much they agreed or disagreed with the statement they had just written about, and provided demographic information.

Results

Exclusions. Applying the same criteria as in our earlier studies, we excluded one participant with incomplete data, four who accused a racially ambiguous suspect of a crime (before the manipulation), and one who took greater than 4.5 *SD*s above the mean time to complete the study. (No IP addresses located participants in non-English-speaking countries). Additionally, 10 participants were excluded for providing an incorrect answer to at least one comprehension-check question (although retaining these participants yielded identical results). Exclusions did not differ significantly by condition, χ^2 (1, N = 76) < 1, p = .42, and left 33 participants in the *insecure identity* condition and 27 in the control condition. Responses on the dependent measure for all remaining participants were within 3 *SD*s of the mean.

Identification of suspects. Participants found the criminal identification task subjectively easy (M = 1.97, SD = 1.14), and 95% of participants always selected the White suspects whose profiles had been designed to look guilty.

Heteroskedasticity. Because we found evidence of heteroskedasticity on all measures reported below, Fs(1, 58) > 5.60, ps < .03 by Levene's test, we report t-tests with adjusted

degrees of freedom (Satterthwaite, 1946). This method for testing differences between two groups is analogous to the *W*-tests used to test differences among multiple groups in Study 4. (Identical results were obtained with traditional *t*-tests).

Manipulation checks. As expected, participants were more concerned that the writing task would make them look bad in the *identity insecurity* condition than in the control condition (*M*s = 3.12 and 1.67, *SD*s = 1.54 and .78, respectively), $t(49.46) = 4.74$, $p < .0001$, $d = 1.35$. The same pattern emerged for how uncomfortable participants felt about the writing task (*M*s = 3.24 and 2.74, *SD*s = 1.64 and 1.26, respectively), but this difference was not significant, $t(57.79) = 1.34$, $p = .19$, $d = .35$. Also as expected, participants expressed greater agreement with the negative statements about teenagers than the negative statements about Blacks (*M*s = 2.00 and -.30, *SD*s = 1.69 and 2.60, respectively), $t(55.33) = 4.13$, $p = .0001$, $d = 1.11$.

Remembering the ambiguous suspects' race. As in Study 5, participants indicated how they recalled the race of the each of the four ambiguous suspects by selecting a morphed photograph, categorizing the suspect's race, and stating how certain they were that he was Black (or White) and how much he resembled a stereotypical Black (or White) individual. As in Study 5, we created a single dependent measure that collapsed across the four ambiguous suspects and combined the photograph, certainty, and stereotypicality scales after standardizing each one ($\alpha = .93$). Higher numbers on this composite measure indicate remembering the suspect as more stereotypically Black.

As hypothesized, participants remembered the ambiguous suspects as looking significantly "Blacker" in the *insecure identity* condition than in the control condition (*M*s = .22

and -.26, *SD*s = 1.9 and .65, respectively), $t(53.45) = 2.13$, $p = .04$, $d = .58$.[4] Comparing these means to the *accuracy motivation* condition from Study 5 ($M = -.43$, $SD = .67$) suggests that participants' estimates were least motivated by accuracy when their identity had been threatened. These findings suggest that fears of seeming racist in the future can lead people to exaggerate the racism of the road not taken.

Individual differences. Results were not moderated by the two measures of prejudice ($M_{mrs} = -.49$, $SD = .78$, possible range: -2 to 2; $M_{atb} = -1.43$, $SD = 1.05$, possible range: -3 to 3; higher numbers indicate greater prejudice), nor by the measure of motivation to respond without prejudice ($M_{ims} = 2.43$, $SD = 1.38$; $M_{ems} = -.84$, $SD = 1.63$; possible range for both measures: 1-9), nor by political orientation ($M_{politics} = 4.16$, $SD = 1.56$, possible range: 1-7), *p*s > .46.

Discussion

Study 6 lends further support to the idea that people invent counterfactual transgressions to assuage concerns about their moral identity. Participants in Study 6 whose identity was *threatened* by the prospect of having to argue in favor of racist statements remembered ambiguous criminal suspects as *more* stereotypically Black. This finding complements the Study 5 result that participants who could *secure* a non-racist identity by rejecting racist statements remembered innocent, racially ambiguous suspects as looking *less* stereotypically Black.

Why might memory distortion secure a non-racist identity? The more stereotypically Black the innocent suspects were, the more clearly racist it would have been to accuse them. Thus, the more stereotypically Black these suspects were remembered as being, the more

[4] Examining each the three items in the composite revealed that this difference was driven primarily by participants' certainty that the ambiguous suspect was Black, $t(55.50) = 1.90$, $p = .06$, $d = .51$, and how stereotypically Black they thought he was, $t(57.46) = 2.76$, $p = .008$, $d = .73$. The means for the photograph scale were in the same direction, but not significantly different, $t(54.26) = 1.28$, $p = .21$, $d = .35$. The condition difference on the composite measure was marginally significant when tested with the mixed model described in Footnote 3, $p < .08$.

participants may have felt that *not* accusing them demonstrated racial egalitarianism. Given how clearly innocent the racially ambiguous suspects were, however, the belief that not accusing them would be diagnostic of egalitarianism seems exaggerated at best.

GENERAL DISCUSSION

Six studies explored how and to what effect people use counterfactual transgressions in securing their moral identity. We found evidence that when people can easily point to an immoral road not taken, they feel that they can act in morally questionable ways without fear of discrediting themselves. A salient racist alternative to their prior behavior made White participants feel that they had evidence of their virtue (Study 1), allowed them to express less racial sensitivity (Study 2), and licensed them to express preferences about employment and allocating money that favored Whites at the expense of Blacks (Study 3). Imagining a counterfactual transgression seems to have made participants feel secure in their morality, and ironically allowed them to act in ways that could seem prejudiced.

The results also showed that the motivation to secure a moral identity leads people to invent immoral alternatives to their past behavior. Specifically, the motivation to obtain evidence of their lack of racism seems to have led White participants to overestimate the number of racist alternatives there had been to their prior behavior (Study 4) and to exaggerate how clearly racist it would have been to perform specific alternative behaviors (Studies 5 and 6). These findings show how people will distort their past in order to make the road not taken seem immoral. Complementing people's tendency to forget information that could call their morality into question (i.e., moral rules that they violated; Shu & Gino, in press; Shu, Gino, & Bazerman, 2011), the present research demonstrated the motivated "remembering" of information that can license one to act in morally questionable ways.

Taken together, the results of these studies are a striking demonstration of how people can feel moral without having performed morally motivated behavior. The findings suggest that by inventing immoral alternatives to their prior behavior that were not actually available, people can address their identity insecurity, and thereby license themselves to act in less virtuous-seeming ways.

Theoretical Mechanisms and Alternative Explanations

How do counterfactual transgressions secure a moral identity? We consider several possible explanations.

Ruling out discrediting identities. The presence of immoral alternatives to one's behavior – whether real or retrospectively invented – may allow one to make charitable inferences about oneself. It is as if such alternatives allow one to think, "Bad people act badly; I did not act badly despite having had an opportunity to do so; therefore, I am not a bad person." In this way, not transgressing may attest to a moral identity by seeming to rule out the possibility that one holds other, discrediting identities such as racist, ingrate, egotist, and so forth. From this perspective, securing a moral identity need not require proof that one is a saint; it may only require evidence that one is not a sinner (cf. Nisan, 1991).

Attributional logic. Perhaps imagining counterfactual transgressions can secure a moral identity by creating the appearance that one resisted strong temptations or situational pressures that would have led most other people to transgress, thus implying that one has a virtuous disposition (cf. Kelley, 1973). Yet little care or effort was required to avoid transgressing in our studies; it seems unlikely that participants would have claimed that they or most others had been tempted to accuse the clearly innocent Black suspects. What people seem willing to accept as

evidence that they are not a sinner is remarkably scanty by conventional attribution logic (Jones & Davis, 1965; Kelley, 1973).

Contrast. A different reason why counterfactual transgressions secure a moral identity may be that they provide a standard against which one's subsequent behavior can be contrasted. There are at least three versions of this contrast effect. First, one's prior neutral behavior (e.g., accusing a clearly guilty suspect of a crime) may take on moral significance when one imagines an immoral alternative (e.g., making a racist accusation instead). Second, a counterfactual transgression may make dubious behaviors seem less problematic in comparison (Markman, Mizoguchi, & McMullen, 2008; Nario-Redmond & Branscombe, 1996), thus increasing one's comfort performing them. Third, counterfactual transgressions may evoke downward social comparisons to less virtuous people than oneself – for example, to racists who would have accused innocent Black suspects (see, e.g., Wills, 1981; Wood, Taylor, & Lichtman, 1985). In this latter view, inventing counterfactual transgressions would represent a strategic attempt to invite such social comparisons (cf. Barkan, Ayal, Gino, & Ariely, in press; O'Brien et al., 2010). These potential mechanisms are not mutually exclusive, and future research should examine when and why each might operate.

Priming. It is difficult to reinterpret the licensing effects observed in Studies 1-3 in terms of the idea that viewing a Black suspect primed a stereotype associating Blacks with crime. First, the Black-stereotypical description of the criminal, which all participants read before viewing the suspects, should have primed this association regardless of experimental condition. Second, it unclear why priming would make participants interpret their selection of a White candidate as evidence of their virtue (Study 1). Third, it is unclear why being primed with a *counterexample* of this stereotype (i.e., a clearly innocent Black suspect) would make

participants willing to express less racial sensitivity and a stronger preference for privileging Whites (Studies 2 and 3). Fourth, participants did not act licensed when they read about the crime and the innocent Black suspect detained by the police, but had no opportunity to accuse this suspect themselves (Study 3). Instead, the data suggest that participants derived a license from the racist behaviors that they could have done but that they did not do.

Moreover, a priming mechanism struggles to account for the results of Studies 4-6. When participants were required to express agreement with racist statements, participants exaggerated the extent to which it would have been possible to make racist judgments on an earlier task (Study 6). Such exaggeration was eliminated when participants saw racist statements but were allowed to express *disagreement* (Studies 4 and 5). These results comport with our reasoning that agreeing with such statements is self-threatening while rejecting such statements secures a non-racist identity. The idea that exposure to the statements primed the concept of racism is insufficient to explain these results.

Individual differences

As in prior work on moral credentials (Monin & Miller, 2001a), the results of the present studies were not significantly moderated by either internal or external motivations to respond without prejudice (Plant & Devine, 1998). Similarly, political orientation – which is thought to be associated with the internalization of anti-prejudiced norms (Nail, et al., 2003) – did not moderate our results even though we recruited participants with a wide range of political orientations in Studies 4-6. These findings may speak to the strength of our experimental situations for eliciting the fear of seeming prejudiced, even among individuals who are not as chronically concerned with avoiding prejudiced behavior.

Our results were also not significantly moderated by racial attitudes as measured by the MRS and ATB. By contrast, prior work suggested that relatively prejudiced individuals are more likely to use non-racist credentials to favor Whites (Effron, et al., 2009, Study 3), and more likely to invent non-racist credentials when a subsequent task could make them seem racist (Effron, 2011, Study 5). Future research is needed to clarify the conditions under which racial attitudes are associated with these moral credentialing effects. Perhaps it is more difficult to make less-prejudiced individuals feel like they need non-racist credentials (e.g., to tempt them to favor Whites), but when these individuals do find themselves needing credentials, they employ the same strategies to obtain them as more-prejudiced individuals do.

Are the Effects of Counterfactual Transgression Merely Self-Presentational?

Several methodological features of the present studies suggest that participants tried to prove their virtue not only to the researchers, but also to themselves. First, it seems likely that participants in Studies 2 and 3 assumed that their responses to the manipulation would be observed by a different audience than the one that would observe their responses to the dependent measure, given that the manipulations and measures were disguised as separate studies and embedded in a series of questionnaires from multiple researchers (see also Monin & Miller, 2001a, Study 3). Second, self-presentational concerns in Studies 4-6 should have been minimized by the fact that participants completed these studies on the Web rather than face-to-face with an experimenter. Third, it seems unlikely that participants would lie to the researchers about the race of the suspects in Studies 4-6, given that the researchers knew which suspects had been presented. These methodological features do not rule out the role of self-presentation definitively, but it seems likely that the use of counterfactual transgressions, like moral

credentials (Monin & Miller, 2001a), are not merely self-presentational (for discussions, see Merritt, et al., 2010; Miller & Effron, 2010).

Implications for Racial Attitudes

Feeling secure in a non-racist identity may sometimes have desirable consequences. Reducing people's concern with feeling or appearing racist may liberate them to have frank, constructive conversations about racial issues, experience less anxiety during interracial interactions, or express legitimate views that could be misinterpreted as prejudiced (e.g., criticizing a public figure who just happens to be Black). Indeed, the dependent measures used in Study 2 provided participants with potentially defensible reasons to favor Whites instead of Blacks (Effron, et al., 2009; Monin & Miller, 2001a). At the same time, our findings raise the concern that people will feel licensed to perform truly prejudiced behaviors by reflecting on (or inventing) even more prejudiced behaviors they could have performed instead.

The present results suggest that it takes surprisingly little to increase people's security in their non-racist identities. One reason may be that many White Americans equate racial bias with overt bigotry and not with the subtler forms of prejudice that social psychologists have documented (Sommers & Norton, 2006). These lay definitions of racial bias set a relatively low bar for what constitutes evidence of one's lack of prejudice (O'Brien, et al., 2010). When a racist is defined only as someone who constantly commits blatant acts of discrimination (e.g., accusing clearly innocent Black suspects of crimes), it is easier to feel that foregoing such acts proves that one is not a racist. We suspect that people can draw on similarly narrow definitions of unethical behavior in other domains to convince themselves that foregoing a single opportunity to lie, cheat, or steal proves that they are not a liar, cheater, or thief, and thus licenses them to commit other ethically questionable acts.

Implications for Moral Behavior

Moral licensing. These findings advance research on moral licensing (for reviews, see Merritt, et al., 2010; Miller & Effron, 2010) by suggesting that doing good deeds is not always necessary for people to feel licensed to deviate from the morally straight and narrow. Merely reflecting on the bad deeds that one did not perform (but could have performed) can ironically increase the likelihood that one will behave in seemingly less virtuous ways in the future. This may help explain why people are willing to license themselves based on morally unremarkable behavior, such as rejecting blatantly sexist statements (Monin & Miller, 2001a), endorsing a popular political candidate who happens to be Black (Effron, et al., 2009), or agreeing to provide help in a hypothetical scenario (Khan & Dhar, 2006). Perhaps such behaviors seem morally significant because one imagines the transgressions that one could have performed instead, such as agreeing with blatantly sexist statements, refusing to endorse the candidate because of his race, or declining even hypothetical help to a person in need. In this way, counterfactual transgressions may allow people to make a moral mountain out of a molehill.

Moral compensation. The findings also speak to how people respond to identity insecurity. An optimistic implication of prior work is that people will act more virtuously when motivated to atone for their prior misdeeds (e.g., Carlsmith & Gross, 1969; Jordan, et al., 2011; Tetlock, et al., 2000). The present research, however, raises the possibility that people may strategically invent counterfactual transgressions in lieu of performing good deeds in order to alleviate moral insecurity caused by their past behavior, the present situation (Studies 4 and 5), or what they plan to do in the future (Study 6). If one can feel virtuous merely by reflecting on distorted memories of the past, why make the effort to perform virtuous behavior?

Conclusion

Counterfactual transgressions provide a potentially nefarious source of flexibility in people's moral lives. When our track records contain a dearth of virtuous behaviors, we can nonetheless take comfort in the knowledge that it would have been possible for us to transgress more than we actually did. And when our track records also lack opportunities to transgress, we can nonetheless distort the past so that such opportunities appear to have existed.

Our identities are defined not only by the paths we have chosen, but also by the paths we believe that we could have chosen instead. These alternative paths may have played no role in determining our choice, or they may not have existed at all, materializing only in retrospect as imagined possibilities. Nonetheless, these roads not taken may dramatically shape the way we evaluate our moral character, and help determine whether we allow ourselves to act on morally questionable impulses. For better or for worse, the life unlived may liberate us to live the life we want.

References

Allport, G. W. (1937). *Personality: A Psychological Interpretation*. Oxford, England: Henry Holt.

Aquino, K., & Reed, A., II. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83*(6), 1423-1440.

Barkan, R., Ayal, S., Gino, F., & Ariely, D. (in press). The pot calling the kettle black: Distancing response to ethical dissonance. *Journal of Experimental Psychology: General*.

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*, 1564-1575.

Brigham, J. C. (1993). College students' racial attitudes. *Journal of Applied Social Psychology, 23*(23), 1933-1967. doi: 10.1111/j.1559-1816.1993.tb01074.x

Carlsmith, J. M., & Gross, A. E. (1969). Some effects of guilt on compliance. *Journal of Personality and Social Psychology, 11*(3), 232-239. doi: 10.1037/h0027039

Conway, P., & Peetz, J. (in press). When does feeling moral actually make you a better person? Conceptual abstraction moderates whether past moral deeds motivate consistency or compensatory behavior. *Personality and Social Psychology Bulletin*.

Crandall, C. S., & Eshleman, A. (2003). A justification-suppression model of the expression and experience of prejudice. *Psychological Bulletin, 129*(3), 414-446.

Crosby, J. R., & Monin, B. (2012). How the opinions of racial minorities influence discrimination judgments. *Manuscript under review*.

Devine, P. G., Monteith, M. J., Zuwerink, J. R., & Elliott, A. J. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology, 60*, 817-830.

Dovidio, J. F., & Gaertner, S. L. (1998). On the nature of contemporary prejudice: The causes, consequences, and challenges of aversive racism. In J. L. Eberhardt & S. T. Fiske (Eds.), *Confronting racism: The problem and the response.* (pp. 3-32). Thousand Oaks, CA US: Sage Publications, Inc.

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin, 23*(3), 316-326.

Dutton, D. G., & Lake, R. A. (1973). Threat of own prejudice and reverse discrimination in interracial situations. *Journal of Personality and Social Psychology, 28*, 94-100.

Dutton, D. G., & Lennox, V. L. (1974). Effect of prior "token" compliance on subsequent interracial behavior. *Journal of Personality and Social Psychology, 29*(1), 65-71.

Effron, D. A. (2011). *The immoral road not taken: Using counterfactual transgressions to secure a virtuous identity and to regulate one's moral behavior.* PhD Doctoral Dissertation, Stanford University, Stanford, CA.

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

Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism.* (pp. 61-89). San Diego, CA US: Academic Press.

Jones, E. E., & Davis, K. E. (1965). A theory of correspondent inferences: From acts to dispositions. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 2, pp. 219-266). New York: Academic Press.

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*, 701-713.

Kelley, H. H. (1973). The process 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.

Mandel, D. R., & Dhami, M. K. (2005). "What I did" versus "what I might have done": Effect of factual versus counterfactual thinking on blame, guilt, and shame in prisoners. *Journal of Experimental Social Psychology, 41*(6), 627-635. doi: 10.1016/j.jesp.2004.08.009

Markman, K. D., Mizoguchi, N., & McMullen, M. N. (2008). "It would have been worse under Saddam:" Implications of counterfactual thinking for beliefs regarding the ethical treatment of prisoners of war. *Journal of Experimental Social Psychology, 44*(3), 650-654. doi: 10.1016/j.jesp.2007.03.005

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

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.

McConahay, J. B., Hardee, B. B., & Batts, V. (1981). Has racism declined in America? It depends on who is asking and what is asked. *Journal of Conflict Resolution, 25*, 563-571.

Merritt, A. C., Effron, D. A., Fein, S., Savitsky, K., Tuller, D. M., & Monin, B. (2012). The strategic pursuit of moral credentials. *Journal of Experimental Social Psychology, 48*, 774-777.

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*, 344-357.

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

Miller, D. T., Visser, P. S., & Staub, B. D. (2005). How surveillance begets perceptions of dishonesty: The case of the counterfactual sinner. *Journal of Personality and Social Psychology, 89*(2), 117-128.

Monin, B., & Miller, D. T. (2001a). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology, 81*(1), 33-43.

Monin, B., & Miller, D. T. (2001b). *Testing mediators of the moral credentials effect*. Unpublished data.

Monin, B., Sawyer, P. J., & Marquez, M. J. (2008). The rejection of moral rebels: Resenting those who do the right thing. *Journal of Personality and Social Psychology, 95*(1), 76-93.

Morpheus Software, L. (2011). Morpheus Photo Morpher (Version 3.15).

Nail, P. R., Harton, H. C., & Decker, B. P. (2003). Political orientation and modern versus aversive racism: Tests of Dovidio and Gaertner's (1998) integrated model. *Journal of Personality and Social Psychology, 84*(4), 754-770.

Nario-Redmond, M. R., & Branscombe, N. R. (1996). It could have been better or it might have been worse: Implications for blame assignment in rape cases. *Basic and Applied Social Psychology, 18*, 347-366.

Niedenthal, P. M., Tangney, J. P., & Gavanski, I. (1994). "If only I weren't" versus "If only I hadn't": Distinguishing shame and guilt in conterfactual thinking. *Journal of Personality and Social Psychology, 67*, 585-595.

Nisan, M. (1991). The moral balance model: Theory and research extending our understanding of moral choice and deviation. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behavior and development* (pp. 213-249). Hillsdale, NJ: Erlbaum.

O'Brien, L. T., Crandall, C. S., Horstman-Reser, A., Warner, R., Alsbrooks, A., & Blodorn, A. (2010). But I'm no bigot: How prejudiced White Americans maintain unprejudiced self-images. *Journal of Applied Social Psychology, 40*(4), 917-946.

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*, 1-25.

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology, 75*, 811-832.

Roese, N. J. (1994). The functional basis of counterfactual thinking. *Journal of Personality and Social Psychology, 66*(5), 805-818. doi: 10.1037/0022-3514.66.5.805

Ross, M., McFarland, C., Conway, M., & Zanna, M. P. (1983). Reciprocal relation between attitudes and behavior recall: Committing people to newly formed attitudes. *Journal of Personality and Social Psychology, 45*(2), 257-267. doi: 10.1037/0022-3514.45.2.257

Ross, M., McFarland, C., & Fletcher, G. J. (1981). The effect of attitude on the recall of personal histories. *Journal of Personality and Social Psychology, 40*(4), 627-634. doi: 10.1037/0022-3514.40.4.627

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

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*, 110-114.

Sears, D. O. (1988). Symbolic racism. In P. A. Katz & D. A. Taylor (Eds.), *Eliminating racism: Profiles in controversy.* (pp. 53-84). New York, NY US: Plenum Press.

Shu, L. L., & Gino, F. (in press). Sweeping dishonesty under the rug: How unethical actions lead to forgetting of moral rules. *Journal of Personality and Social Psychology*.

Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality and Social Psychology Bulletin, 37*, 330-349.

Sommers, S. R., & Norton, M. I. (2006). Lay theories about white racists: What constitutes racism (and what doesn't). *Group Processes and Intergroup Relations, 9*(1), 117-138.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics*. Boston: Pearson Education.

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78*(5), 853-870.

Welch, B. L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika, 38*, 330-336.

White, K., & Lehman, D. R. (2005). Looking on the bright side: Downward counterfactual thinking in response to negative life events. *Personality and Social Psychology Bulletin, 31*(10), 1413-1424.

Wilcox, R. R. (2003). *Applying Contemporary Statistical Techniques*. Los Angeles: Academic Press.

Wills, T. A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin, 90*(2), 245-271. doi: 10.1037/0033-2909.90.2.245

Wood, J. V., Taylor, S. E., & Lichtman, R. R. (1985). Social comparison in adjustment to breast cancer. *Journal of Personality and Social Psychology, 49*(5), 1169-1183. doi: 10.1037/0022-3514.49.5.1169

Zhong, C.-B., Ku, G., Lount, R. B., & Murnighan, J. K. (2010). Compensatory ethics. *Journal of Business Ethics, 92*(3), 323-339. doi: 10.1007/s10551-009-0161-6

Zhong, C.-B., Liljenquist, K., & Cain, D. M. (2009). Moral self-regulation: Licensing and compensation. In D. De Cremer (Ed.), *Psychological Perspectives on Ethical Behavior and Decision Making* (pp. 75-89). Charlotte, NC: Information Age Publishing.

Figures

Figure 1. Study 2: Mean number of statements labeled as racist (± *SE*), out of 5, by condition

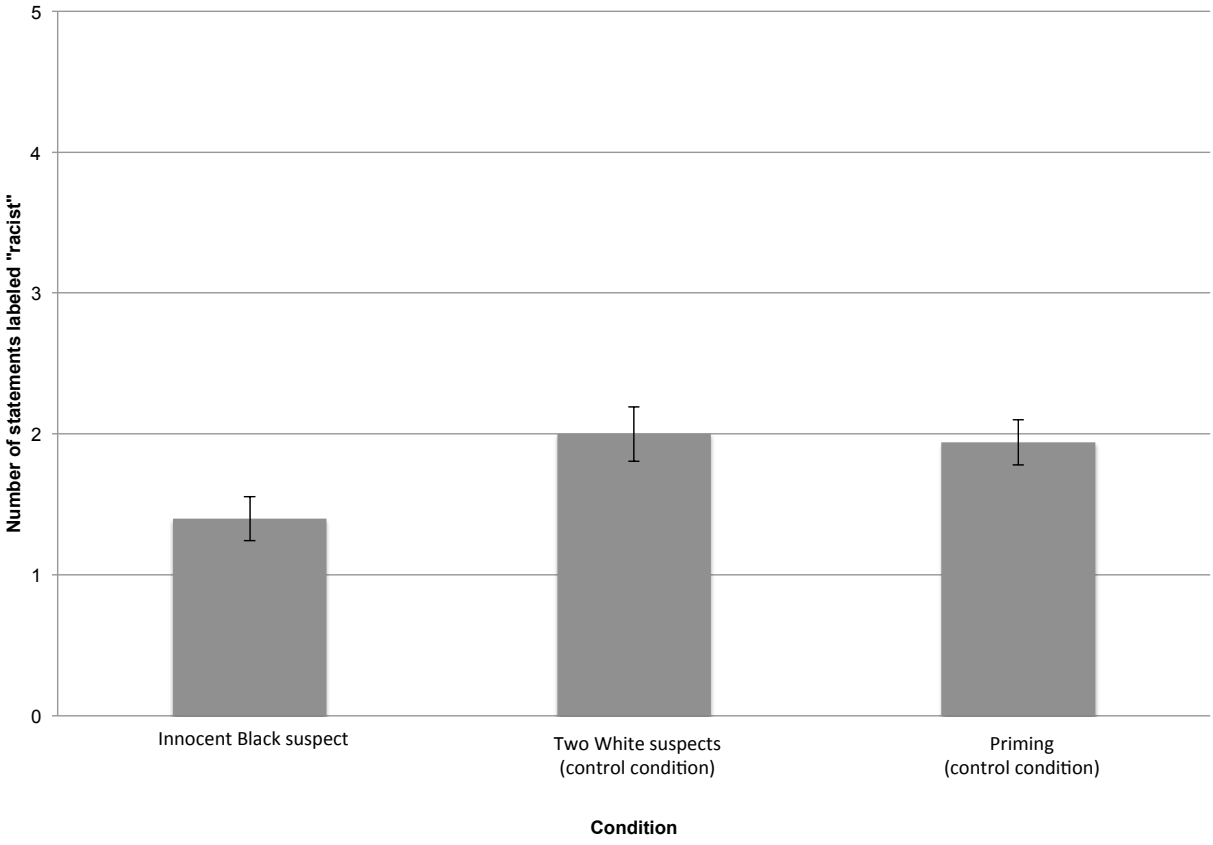
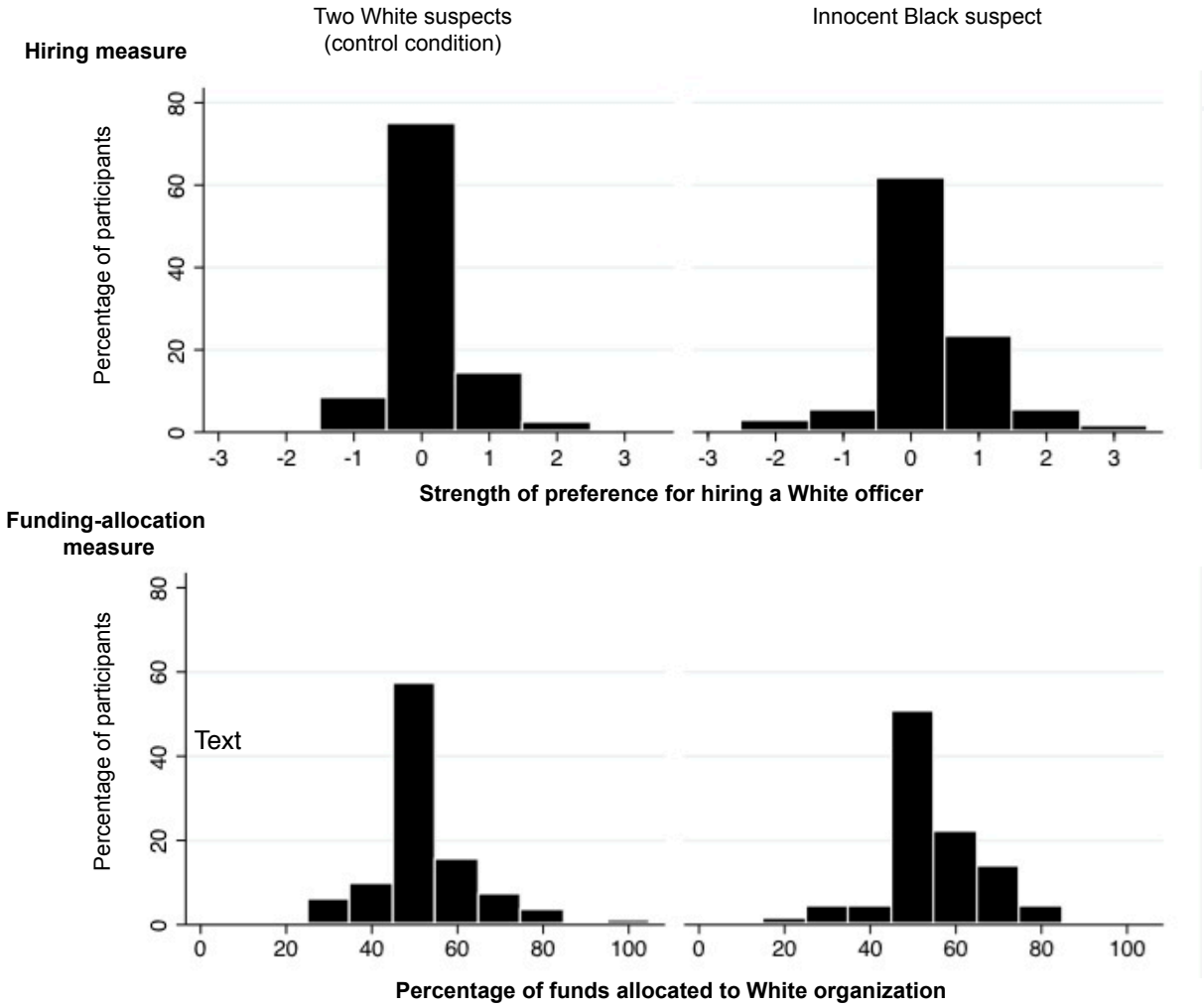


Figure 2. Study 3: Histograms for measures of favoring Whites vs. Blacks, by condition



Note: On the hiring measure (top panel), negative numbers indicate a preference for hiring a Black officer, 0 indicates no racial preference, and positive numbers indicate a preference for hiring a White officer. On the funding-allocation measure (bottom panel), 50% indicates an equal division of funds between the White and Black organizations. Outlying responses shown in the figure were recoded for analysis (see main text).

Figure 3. One of 16 crime stimuli used in Study 4.

The Crime

Mugging a jogger in the park

The Lineup

	Suspect #1	Suspect #2	Suspect #3
Pic			
Alibi	Walking alone elsewhere in the park (unconfirmed)	At home alone (unconfirmed)	Having dinner with friends (confirmed by friends)
Prior record	Assault	None	Downloading music illegally
Behavior when detained	Became angry and yelled at officer	Cooperated with officer	Cooperated with officer
Other evidence	Found with concealed weapon (knife)	Nothing of note	Nothing of note

Click the button under the suspect who you personally think is **most likely** to have committed the crime.

Note: The images shown are samples that closely resemble the stimuli presented to participants. Sample images are courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition, Carnegie Mellon University, http://www.tarrlab.org/. The images shown to participants can be requested from Jennifer Eberhardt, http://www.stanford.edu/~eberhard/cgi-bin/wordpress/.

Figure 4. Study 4: Mean number of foregone opportunities for racism "remembered" by participants (± *SE*), across conditions

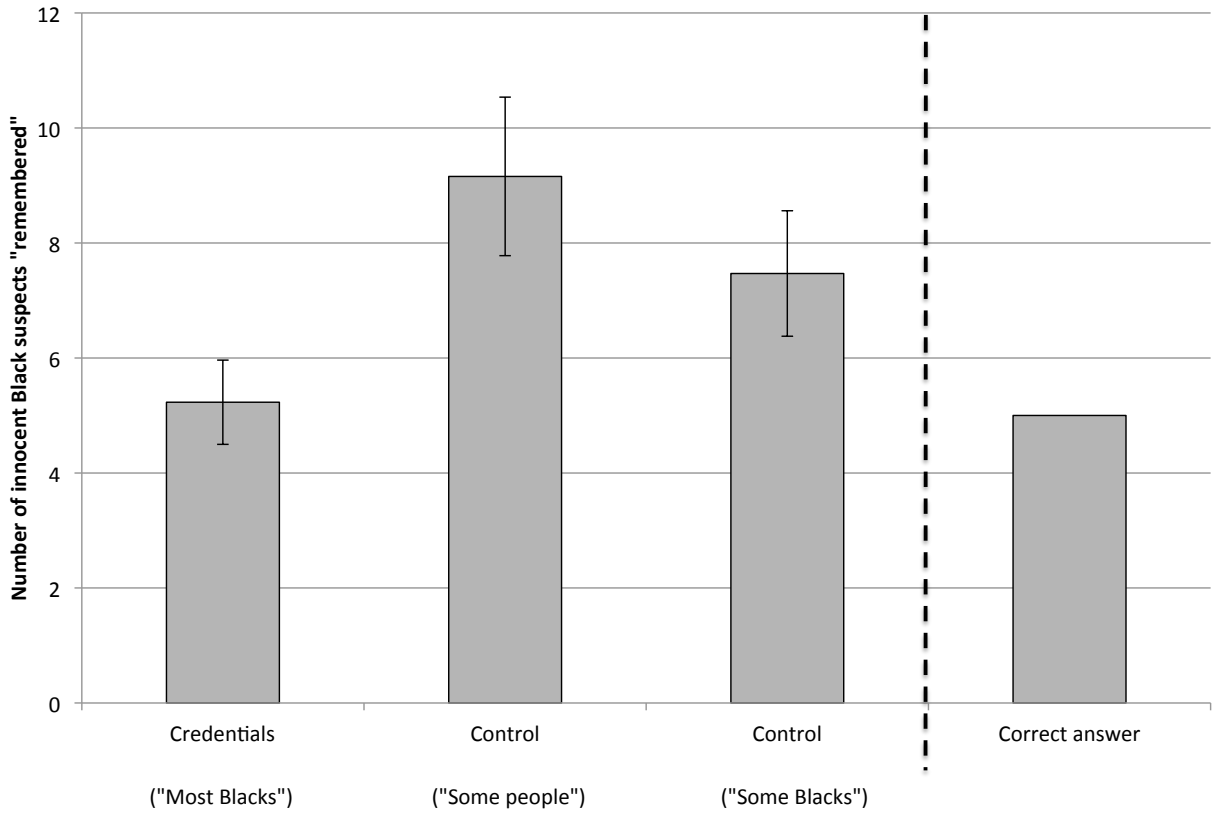
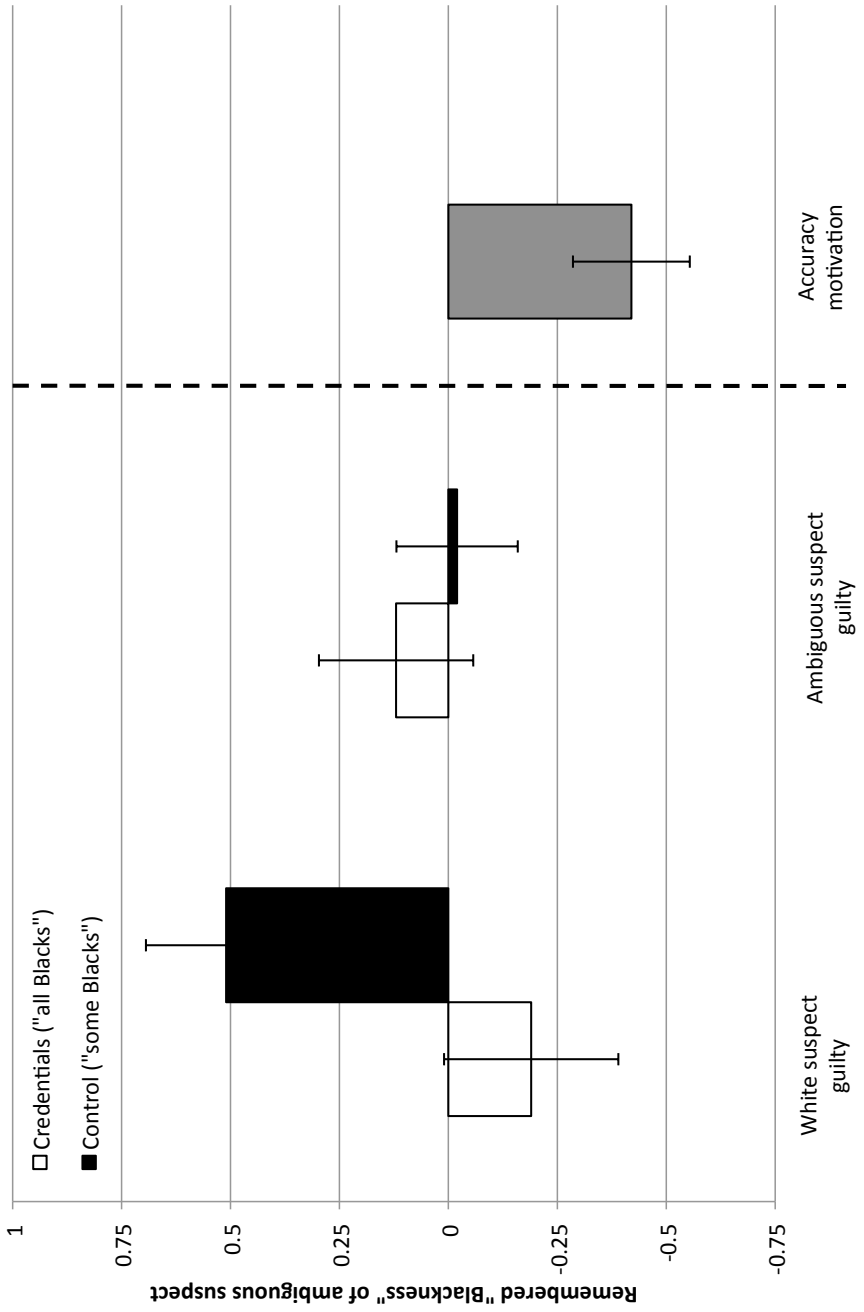


Figure 5. Study 5: Remembered "Blackness" of racially ambiguous suspect (mean ± *SE*), by manipulations of participants' non-racist credentials and suspects' guilt



Note: The y-axis displays a standardized composite of three items measuring how stereotypically Black vs. White participants remembered the racially ambiguous suspects looking. Higher numbers indicate remembering them as looking more stereotypically Black.