The Rejection of Moral Rebels: Resenting Those Who Do the Right Thing

Benoît Monin
Stanford University

Pamela J. Sawyer
University of California, Santa Barbara

Matthew J. Marquez
New York

Four studies document the rejection of moral rebels. In Study 1, participants who made a counterattitudinal speech disliked a person who refused on principle to do so, but uninvolved observers preferred this rebel to an obedient other. In Study 2, participants taking part in a racist task disliked a rebel who refused to go along, but mere observers did not. This rejection was mediated by the perception that rebels would reject obedient participants (Study 3), but did not occur when participants described an important trait or value beforehand (Study 4). Together, these studies suggest that rebels are resented when their implicit reproach threatens the positive self-image of individuals who did not rebel.

Keywords: moral rebels, whistleblowers, implicit moral reproach, resentment, self-affirmation

Rebels hold a special place in social psychology. In a field that often underscores the evils (and the power) of conformity and obedience, people willing to go against the grain in the name of their principles are presented as the lone exceptions that restore our faith in human nature. When we report classic findings showing that a majority of people agree to hurt an innocent victim (Milgram, 1974), fail to help a person in need (Latané & Darley, 1970), or simply conform to an erroneous pronouncement (Asch, 1956), we find solace in the fact that a minority of respondents hold their own and do the honorable thing. Accounts of the notorious 1968 My Lai massacre (My Lai Massacre, 2006) contrast the destructive obedience of William Calley and his men with the decency of Hugh Thompson Jr., the helicopter pilot who stopped the massacre by standing up against Calley. We celebrate men like Frank Serpico, the New York Police Department (NYPD) police officer who took a stand against corruption at a time when it was rampant in the force (Maas, 1973). And one would hope that Joseph Darby, the military policeman who turned in the Abu Ghraib picture CD to authorities, will similarly go down in history as a hero who prevented further abuse.

Disliking Moral Rebels

It may therefore come as a surprise that these inspiring rebels all suffered intense backlash from their peers. Hugh Thompson was

shunned for years by fellow soldiers and received numerous death threats (My Lai Massacre, 2006); Frank Serpico was shot in the face in a suspected setup by fellow NYPD officers, prompting him to leave the country for a decade (Maas, 1973); and Joseph Darby had to be taken into protective military custody at an undisclosed location after receiving various threats from former colleagues (Rosin, 2004). The violence of this backlash against whistleblowers and rebels is surprising precisely because the exact same behavior draws admiration and respect from observers not directly involved in the situation—and also because this rejection does not just come from peers who stand to suffer from the rebellion, but also from peers who merely failed to report or oppose the abuse.

A similarly puzzling reversal is captured in a variation of Milgram's (1965) obedience paradigm. Whereas readers of the original studies typically applaud the third of participants who refused to shock the victim all the way as strong, reliable, and altruistic moral exemplars, obedient participants had a very different opinion. When naïve participants were paired with confederates instructed to stand up against the experimenter (Milgram, 1965, Study 2), those participants who kept shocking the victim throughout the procedure were quick to put down rebels at debriefing. Instead of seeing rebels as strong, they thought, for example, that rebels were "just being ridiculous" and that they "lost all control of themselves" (p. 132). Instead of seeing rebels as reliable, they thought that they should not have quit ("They came here for an experiment, and I think they should have stuck with it," p. 132). Instead of seeing rebels as altruistic, they saw them as ignoring the needs of the experimenter ("If [. . .] I did the same, I don't know how many months and days you'd have to continue before you got done," p. 132). Just as in the case of the moral rebels depicted in news reports, the exact same behavior seems to be sanctified by some, and despised by others, depending on their involvement in the situation.

The goal of this article was to document and understand this backlash against moral rebels. We define *moral rebels* as individuals who take a principled stand against the status quo, who refuse

to comply, stay silent, or simply go along when this would require that they compromise their values. We predict that their rebellion will be inspiring to uninvolved observers (e.g., civilians hearing of Joseph Darby, or readers of Milgram's obedience studies) but threatening to people in the situation (e.g., coworkers of Joseph Darby who did not report similar abuses, obedient participants in Milgram's studies), whose own behavior is implicitly called into question and who will dislike rebels as a result. Although we started with dramatic examples, this threat can be observed in everyday settings. When a doctor decides to refuse lavish gifts from drug companies, we would predict that this gesture would inspire more respect from residents than from fellow doctors who have accepted such perks in the past and may perceive this refusal as an implicit indictment. When a student refuses, on principle, to download pirated music from the Internet, we would predict that this choice makes the student more likable to peers who do not own computers and have never had the opportunity to download a song than to peers who routinely download pirated music without (until now) a second thought.

The Root of Resentment: Imagined Reproach Is a Threat to the Self

Where does this backlash come from? In the examples above, personal involvement seems to be an important moderator of the reaction to moral rebels. Rebels may think that they are only taking a stand against the status quo, but bystanders who did not take that stand can take this rebellion as a personal threat. This suggests that the root of resentment may be that the rebel's choice implicitly condemns the perceiver's own behavior and that this potential reproach shakes the perceiver's confidence in being a good, moral person (their sense of "moral and adaptive adequacy"; Steele, 1988, p. 262; see also Sherman & Cohen, 2006). Milgram (1965) described how "the reaction of the defiant confederate defines the act of shocking the victim as improper" and how "each additional shock administered by the naïve subject now carries with it a measure of social disapproval from the two [defiant] confederates" (p. 133). We propose that the rebel does not even need to be in the room, nor does he or she need to know of the actor's behavior, for his or her stance to be threatening; his or her gesture alone stands as a claim "defining the act as improper." Nonrebels basically assume that rebels espouse the indictment of passivity attributed to Black Panther activist Eldridge Cleaver that "you're either part of the solution, or you're part of the problem" (Cleaver, 1969, p. xxxii)—and no one likes to be called part of the problem. By taking a moral stand, rebels imply that it is wrong for anyone else not to do the same because moral dictates are by definition universal (Frankena, 1973; Turiel, 1983). By claiming the moral high ground, rebels are effectively calling everything else the low road.

Moral reproach, even imagined, can be extremely threatening to individuals' sense of adequacy. Sabini and Silver (1982) noted how sensitive individuals are to moral reproach because of the centrality of morality in most people's self-concept (Allison, Messick, & Goethals, 1989) and because they are aware of the social stigma that accompanies having one's morality questioned (Park, Ybarra, & Stanik, 2007). It may therefore not be surprising that actors have little fondness for someone whose behavior amounts to a wag of the finger at their own. Other people's moral claims, and the perception that they could look down on us for our choices,

might shake our confidence in our own adequacy. Only the most self-confident of individuals would welcome such an implicit challenge with equanimity. If that is the case, then manipulations that comfort individuals that they are good, able people (self-affirmation; Steele, 1988) and that have been shown to reduce the threat of superior others (Spencer, Fein, & Lomore, 2001) might reduce the need to put down moral rebels to protect the self.

Two things need to be stressed about this imagined reproach: First, we propose that it is not necessary for moral rebels to condemn explicitly those who did not rebel. The very fact of taking a moral stance should be perceived as an implied reproach against (and implicit rejection of) those not making the same choice. Second, virtual reproach may be enough to trigger resentment, with no need for the rebel to ever actually know of the conformist's behavior. A newspaper article or Web site castigating our way of life can be irritating even if the authors never met us. And by extension, a moral rebel seen on television may irk a viewer whose behavior is implicitly called in question, even if there is no chance that the rebel will ever meet or for that matter form a judgment about the viewer. It is the fact that the rebel most likely would reproach the viewer (and the ensuing self-threat) that triggers the rejection, whereas an unvested bystander might embrace the rebel.

Righteous Indignation or Self-Righteous Whininess?

Now that we have articulated what we believe to be the underlying causes of the rejection, let us elaborate on its content: What personality dimensions are rebels put down on? A striking feature of our opening examples is that the exact same behavior can be constructed in such different ways depending on the perceiver's involvement. One reason why it is easy to demote moral behavior might precisely be because individuals hold multiple moral prototypes, so they can opportunistically emphasize the aspect of morality that best preserves their self-image. Walker and Hennig (2004) identified just, brave, and caring as three distinct moral prototypes, corresponding to different profiles in the two-dimensional space defined by agency/dominance and communion/nurturance. As in the Milgram (1965) example above, observers may see rebels as the embodiment of righteous agency (closer to Walker and Hennig's "brave" prototype) for standing up against an unjust situation, whereas threatened actors may deny that it took any strength of character to rebel and instead define the rebel's stance in terms of lack of communion (further from the "caring" prototype). Besides overall social attraction, the studies presented here therefore strive to identify the dimensions of interpersonal judgment used by individuals when encountering moral rebels, paying particular attention to the dimensions of communion and agency.

Testing the Model

We now describe more formally the hypotheses being proposed and outline the studies that we present to test them.

Hypotheses

To facilitate the evaluation of the claims presented here, three hypotheses (and six related predictions) can be formulated.

Hypothesis 1: The Perversity of Obedience

The simple fact of obeying in a problematic situation should make individuals like a rebel less (relative to an obedient other). Thus, obedient actors not only go along with a problematic situation but also perversely become its guardian by putting down those who resist it. This predicted interaction could be broken down into two simple effects:

Prediction 1a: Rejection by actors. Actors should like rebels less than they like obedient others. They should not give rebels any credit for their rebellious behavior (no agency or morality effect) and justify rejecting rebels by casting them as not very nice people (low communion). Studies 1, 2, 3, and 4 tested this prediction.

Prediction 1b: Attraction by observers. Observers, however, should like rebels more than they like obedient others. They should appreciate the rebels' strength of character (higher agency) and moral righteousness (higher ratings of morality), regardless of how nice rebels are seen to be (no communion effect). Studies 1, 2, and 3 tested this prediction.

Hypothesis 2: Preemptive Rejection

The first step in triggering resentment is the perception that rebels look down on those who did not rebel, and they would reject them if they met them. Rejection would therefore be a preemptive strike, to put down someone who is in a position to put down the actor. From this hypothesis, we can make two concrete predictions.

Prediction 2a: Imagined rejection. Actors should expect to be liked and respected less by rebels than by obedient others. Studies 3 and 4 tested this prediction.

Prediction 2b: mediation. Rejection of rebels should be a function of how much actors imagine that they would be rejected by rebels. Studies 3 and 4 tested this prediction.

Hypothesis 3: Self-Threat

The second step in explaining resentment is that reproach, even imagined, shakes actors' overall sense of self-worth. The rejection of rebels would be an attempt to deny this vulnerability and to preserve one's sense of being a good person. If this is true, then individuals who have been secured in their moral and adaptive adequacy, that is, self-affirmed (Steele, 1988), should show less need to reject rebels or deny the implications of their stance.

Prediction 3a: Self-affirmation opens the heart. Self-affirmed actors should not feel a need to reject rebels as much as individuals less secure in their sense of self-worth, even if they still believe that rebels would dislike them. Study 4 tested this prediction.

Prediction 3b: Self-affirmation opens the eyes. Not needing to deny the rebels' gesture to protect a fragile sense of worth, self-affirmed actors should be able to recognize its value and draw appropriate conclusions about their own behavior. Study 4 tested this prediction.

Overview of the Present Studies

We present four studies testing the hypothesis that a moral rebel is rejected when others are personally threatened by the rebellion as an implied condemnation of their own conformity. In all four studies, participants in the focal experimental condition agree to go along with a problematic request from the experimenter (speaking against their beliefs in Study 1, playing a racist game in Studies 2–4), only to discover after the fact that another participant (actually a confederate) refused, on principled grounds, to comply with the experimenter. Control participants rate an obedient confederate, or are not asked to perform the problematic task beforehand, or both. Furthermore, Study 3 assessed the role of imagined rejection as a factor in resentment by testing whether rejection is mediated by the fear of being rejected by the rebel, and Study 4 tested the involvement of self-threat by testing whether buttressing one's sense of self-worth and integrity through self-affirmation (Sherman & Cohen, 2006; Steele, 1988) eliminates the rejection of moral rebels.

Study 1: Writing a Problematic Speech

We used a variation of the induced compliance paradigm (Galinsky, Stone, & Cooper, 2000; Zanna & Cooper, 1974) in Study 1 to first investigate the rejection of moral rebels. In this classic procedure of the cognitive dissonance literature, participants write a speech that goes against their own attitude, and when they perceive that they could have easily refused to write (high-choice condition), they tend to change their attitude in line with their speech, presumably to reduce the dissonance created by writing it (Festinger, 1957). Participants who are simply told to write the speech with no room for refusal (low-choice condition) do not typically change their attitude as much.

We focused on the latter condition, reasoning that low-choice participants were similar to obedient observers of moral rebels: They are unquestionably going along with a behavior that goes against their own attitude and that, they believe, will have concrete negative consequences (Cooper & Fazio, 1984; Scher & Cooper, 1989). They have ample motive to stand up for their own attitude, and yet they do not. And we know from the results of high-choice conditions that without the safety blanket of the low-choice manipulation, they would feel a measure of discomfort associated with misrepresenting their true self. So how would they react to the news that another participant in the same condition refused to go along? As in the real-world examples presented above, we predicted that disinterested raters (*observers*) would like and admire a rebel standing up for his or her opinion more than an obedient other (Prediction 1b) but that participants who have themselves gone along with the problematic behavior (*actors*) by agreeing to write a counterattitudinal speech in a low-choice condition would instead reject the rebel (Prediction 1a).

Beyond this global rejection (which we measured with liking and respect items), we wanted to document the nature of the impressions formed about the rebels by actors and observers, and in particular to document the role of agency and communion in the perception of rebels. In the introduction, we posited that observers would see rebels as strong individuals (higher agency), but not necessarily nicer than obedient others (no effect on communion), in line with Prediction 1b, whereas actors would definitely see rebels as not very nice (lower communion) and not give them credit for being strong either (no effect on agency), in line with Prediction 1a. Compared with obedient others, rebels should thus be rated higher on agency by observers (Prediction 1b), and lower on communion by actors (Prediction 1a).

Method

Participants and Design

Seventy undergraduates at a large U.S. private university (24 men, 35 women, 11 unreported) took part in this experiment in a laboratory setting in exchange for $10. Thirty-five first completed the *actor* version of the experiment, and 35 others were recruited 1 month later from the same population to complete the *observer* version.[1] In both versions of the study, participants were randomly assigned to hearing a *rebel* or an *obedient* other, resulting in a 2 (actor vs. observer) × 2 (rebel vs. obedient) between-subjects design. The experimenter was either a White man or a White woman.

Procedure

Participants in the actor version took part in the low-choice version of an induced compliance paradigm (Zanna & Cooper, 1974). Under the guise of studying the relationship between perception of personality and cogency of arguments, the experimenter told participants to make a speech in favor of eliminating the reading week (a class-free period preceding final examinations), a proposal that we knew was very unpopular in our participant population. To instantiate foreseeable aversive consequences (Cooper & Fazio, 1984), we told participants that the tapes would go to the "Undergraduate Committee on Curriculum," which allegedly funded the project and would probably use the arguments when deciding on the reading week. Participants were given a few minutes to prepare their speech, recorded it on their own, completed some questionnaires, and were then introduced to the "personality perception" part of the experiment. In this section, they listened to a tape allegedly recorded by a previous participant in the same setting and used the scales provided (see the *Measures* section below) to indicate what they thought of the other person. The tape that they listened to contained the rebellion manipulation—in the obedient condition, the other complied, and in the rebel condition, he or she rebelled (see the Appendix for scripts of tapes).

In the observer version, participants were given a detailed written description of the instructions (allegedly) received by the recorded speakers. These instructions matched the ones actually used with actors in the induced compliance paradigm. They then listened to the same audiotapes as actors did and rated the speaker, without making a speech of their own. In all conditions, the gender of the speaker was matched with that of the participant.

Measures

Under the guise of investigating the relationship between personality and cogency of arguments, participants rated the speaker on 14 7-point bipolar scales anchored on *stupid–intelligent, weak–strong, insecure–confident, passive–active, cruel–kind, awful–nice, cold–warm, dishonest–honest, unfair–fair, unpleasant–pleasant, dependent–independent, stingy–generous, immature–mature,* and *low self-esteem–high self-esteem.* Participants then indicated, on 7-point scales, ranging from −3 (*dislike very much*) to + 3 (*like very much*), how much they would like to work on a class project with the speaker, how much they would like the speaker as a friend, and how much they would like the speaker as a roommate. Then, they listed three personality traits that came to
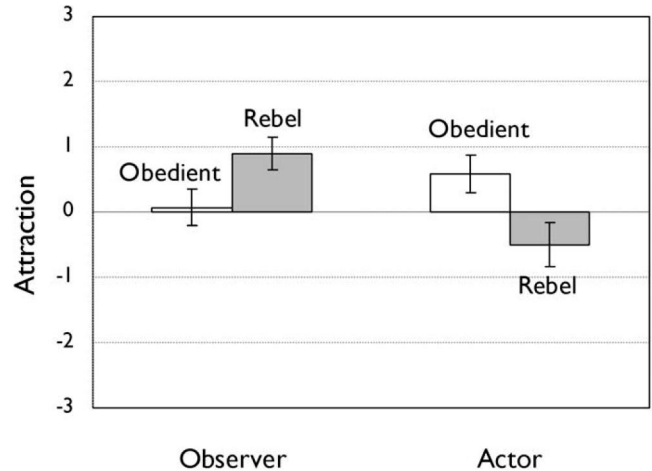

Figure 1. Attraction means (± 1 *SE*) by rebellion condition and role in Study 1. Whereas observers prefer a rebel to an obedient other, actors prefer an obedient other, going as far as disliking the moral rebel.

their mind to describe the speaker, reported how much choice they thought they had in making the speech that they gave (actors only), indicated how much they respected the person on the tape on a 7-point scale ranging from −3 (*despise a great deal*) to +3 (*respect a great deal*), and finally estimated how the speaker felt about eliminating the reading week, on a scale ranging from 1 (*strongly disagree*) to 11 (*strongly agree*).

Results

Suspicion

Ten participants (out of 70) expressed suspicion at debriefing—not uncommon with this type of procedure (see Galinsky et al., 2000). We conducted the analyses below with the 60 nonsuspicious participants first, but results and significance levels for the main analyses were the same with all 70 participants.

Attraction

We created an index of attraction by averaging scores on liking as a friend, as a roommate, on a project, and respect (Cronbach's α = .80) and conducted a Rebellion (rebel, obedient) × Role (actor, observer) analysis of variance (ANOVA) on this index. As predicted, we found a significant interaction between role and rebellion, $F(1, 56) = 11.00$, $p = .002$, $MSE = 1.20$, partial $\eta^2 = .16$, due to the fact that observers preferred the rebel ($M = 0.90$, $SD = 0.87$) to the obedient target ($M = 0.07$, $SD = 1.51$), $t(56) = 2.22$, $p = .03$, $d = -0.72$, whereas actors preferred the obedient other ($M = 0.59$, $SD = 0.70$) to the rebel ($M = -0.50$, $SD = 1.21$), $t(56) = 2.46$, $p = .02$, $d = 1.19$ (see Figure 1). Neither of the main effects was significant (both $ps > .13$, partial $\eta^2 < .05$).

---

[1] Participants were not randomly assigned to the actor or observer group, but all were drawn from the same population, and we had no reason to suspect any systematic preexisting differences between the two groups. Studies 2–4 remedy this issue by randomly assigning participants to all conditions.

Agency and Communion

We first conducted a principal axis factor analysis with a Promax rotation (as recommended by Russell, 2002) on the trait ratings, which suggested a two-factor solution (using the scree plot method) capturing over 62% of the variance. We created two factors by averaging traits with loadings higher than .5 on one and only one factor, yielding a nonoverlapping solution that included most traits (but excluded fair, which loaded less than .5 on both factors). The first factor (agency) averaged independent, strong, confident, active, high self-esteem, honest, intelligent, and mature (Cronbach's $\alpha = .93$), whereas the second factor (communion) averaged pleasant, generous, warm, kind, and nice (Cronbach's $\alpha = .86$, and $r = .18$, *ns*, between the two aggregates).

The Role × Rebellion ANOVA conducted on agency revealed a significant main effect for role, $F(1, 56) = 5.99$, $p = .02$, $MSE = 1.23$, partial $\eta^2 = .10$, qualified by a significant interaction, $F(1, 56) = 5.43$, $p < .05$, partial $\eta^2 = .09$. As predicted, observers saw rebels as more agentic than obedient others, $t(56) = 3.69$, $p < 001$, whereas actors did not, $t(56) = .08$, *ns*. The same analysis on communion revealed a marginal main effect for rebellion, $F(1, 56) = 3.80$, $p < .06$, $MSE = .71$, partial $\eta^2 = .06$, but no significant interaction, $F(1, 56) = 1.9$, $p = .18$, partial $\eta^2 = .03$, despite the fact that actors rated rebels lower on communion than obedient others, $t(56) = 2.17$, $p = .03$, whereas observers showed no such difference, $t(56) = 0.45$, *ns*.

Other Variables

We had asked actors to rate how much choice they thought they had in making the speech on a 9-point scale ranging from 1 (*not at all*) to 9 (*a lot*). Actors who saw a rebel reported feeling more freedom ($M = 3.91$, $SD = 2.70$) than actors who saw an obedient other ($M = 1.85$, $SD = 1.34$), $t(22) = -2.43$, $p = .02$, but this score was not correlated with the attraction composite, $r(22) = .12$, *ns*. We also asked participants how the other participant felt about eliminating reading week and submitted this variable to a Role × Rebellion ANOVA. Not surprisingly, participants believed that the rebel was more against it than the obedient other, $F(1, 54) = 80.50$, $p < .001$, $MSE = 2.97$, but neither the main effect of role ($p > .11$) nor the interaction ($p > .38$) was significant.

Discussion

Study 1 provides initial support for our perversity of obedience hypothesis (Hypothesis 1), that the same act of rebellion can be liked by some and rejected by others, depending on whether the rebel's behavior can be seen as an implicit indictment of the judges' own actions. Considered in the abstract and from a disinterested standpoint, rebels were liked better and even seen as more agentic than obedient others (supporting Prediction 1b), but when participants had already engaged in the problematic task that rebels were standing up against (writing a speech supporting a widely unpopular policy), rebels were now liked significantly less than obedient others (supporting Prediction 1a). As posited above, this effect was obtained even though rebels never explicitly reproached the participants, nor would they ever concretely be in a position to do so, having left the laboratory before the participants arrived.

Yet, the exact same behavior led to liking or rejection, solely as a function of participants' involvement in the problematic situation.

Study 2: Going Along With a Racist Task

Study 2 was designed to provide a second demonstration of the rejection of moral rebels, in a less ambiguously moral domain (racism), and using the minimal case in which the problematic compliance does not entail writing a tedious speech, but instead just pointing the finger at an obvious suspect. We replaced the morally problematic behavior of misrepresenting one's views (Study 1) with going along with a racist task. The stimuli in this task reflected stereotypes about African Americans as criminals, and although bowing to the demands of the task did not per se reflect prejudice on the part of the participant, the moral rebel in Study 2 questioned whether it was even appropriate to take part in a situation that contains offensive elements. As in Study 1, we predicted that observers should like this stance more than obedience (Prediction 1b), whereas actors who complied with the task should prefer obedience (Prediction 1a).

Rather than delve into traditional domains of conservative morality, we used prejudice, an issue known to elicit compunction (see Devine, Monteith, Zuwerink, & Elliot, 1991) in an otherwise morally relativistic participant population. Students receive ample warnings from the day they arrive on campus about the ills of prejudice and discrimination, raising strong motivations not to be or be seen as prejudiced (Plant & Devine, 1998). Furthermore, the contemporary moral heroes that our participants have been exposed to even before graduating from high school are often involved in the fight against prejudice (e.g., Martin Luther King Jr., Gandhi, Nelson Mandela, Oskar Schindler), thus reinforcing the link between prejudice and morality and the exemplarity of standing up against it.

Individuals who confront prejudice may be modern moral heroes, but they are typically less appealing to the targets of their invectives (Czopp, Monteith, & Mark, 2006). In Study 2, the moral rebel never directly confronted participants, but we varied his or her self-relevance by manipulating whether participants witnessed his or her refusal before or after they had complied with the racist task. In line with our perversity of obedience hypothesis (Hypothesis 1), we predicted that the same behavior, standing up against prejudice, would inspire liking or rejection, depending on this manipulation of threat.

Method

Participants

Fifty-six undergraduates at a large U.S. private university (13 Whites, 17 Asians, 2 Mexicans, 2 East Asians, 1 Pacific Islander, 1 Arab, 20 unreported[2]; 19 women, 37 men) were approached by an Hispanic male in their dormitory during a summer session and agreed to complete a short questionnaire.

Procedure and Design

Participants were told that the study was about perceptions of personality and that they would need to judge the personality of a

---

[2] Due to experimenter oversight, ethnicity was not recorded for the first 20 participants. However, throughout the procedure, the experimenter only approached students who did not appear to be African American.

previous participant solely on the basis of his or her responses to a questionnaire. They were randomly assigned to either the observer or the actor condition, which was essentially an order manipulation. In the observer condition, they first rated the alleged other participant (the "target") and then completed the same task themselves. The experimenter handed them a blank rating sheet and shuffled through a stack of completed questionnaires (the "police decision task," see below), apparently picking one randomly for them to rate. In reality, all completed questionnaires were prepared to manipulate rebellion by the target. After finishing the ratings, observers were given a blank version of the police decision task and asked to complete it themselves. In contrast, participants in the actor condition were first given the blank decision task to complete, and only after they had turned in their own sheet did they receive a survey allegedly completed by the target, along with a rating sheet. Within each of the role conditions (actor vs. observer), participants were randomly assigned to one of two rebellion conditions (obedient vs. rebel), yielding four cells.

Stimuli

The police decision task was presented on a single sheet with three pictures. It started with "Imagine that a burglary has happened in a neighborhood, and the police have apprehended three suspects. Below are brief descriptions of the three suspects. Please consider these carefully and indicate who you think is most likely to be guilty." Below these instructions were three photographs, each accompanied by some information (name, alibi, previous record, possessions when apprehended, action when apprehended, occupation), presented in tabular format. Under this information, the instructions went on, "Imagine that you are the detective in charge of this case. Please circle the face of the person who you think is most likely to have committed the burglary. In the space below, indicate the reasons for your suspicion," followed by an empty box with eight blank lines.

The information provided in the table was designed to incriminate the third suspect, Steven Jones: He had no alibi, he had a previous record, he was carrying cash and a screwdriver, and he was unemployed. Steven Jones was also the only African American in the display; the other two suspects were White. Participants encountered this task twice: once blank for them to complete and once filled out to instantiate the rebellion manipulation.

In the obedient condition, the African American face was circled, and the handwriting in the box said, "I think Steven Jones did it because 1) He's got no real alibi, 2) He's done it before, and 3) He's carrying a lot of cash, especially for someone without a job. The screwdriver might have helped to break open a door, etc." In the rebel condition, no face was circled, and the box said, "I refuse to make a choice here—this task is obviously biased. . . . Offensive to make black man the obvious suspect. I refuse to play this game."

Measures

Participants formed an impression of the alleged previous participant and reported this on a personality rating form similar to the one used in Study 1, except that the order of attraction and personality traits was reversed. Participants first indicated on 11-point scales ranging from −5 (*dislike very much*) to +5 (*like very much*) how much they would like the other person to work on a

project, as a friend, and as a roommate. They indicated how much they respected the other person on an 11-point scale ranging from −5 (*despise very much*) to +5 (*respect very much*). In an open-ended format, they were asked to indicate the personality trait that they thought best described the other participant. Finally, they rated the other's personality on 16 7-point semantic differentials ranging from −3 to +3 and anchored at *stupid–intelligent, weak–strong, unpleasant–pleasant, insecure–confident, immature–mature, passive–active, cruel–kind, awful–nice, cold–warm, dishonest–honest, unfair–fair, immoral–moral, dependent–independent, selfish–generous, low self-esteem–high self-esteem,* and *rude–polite.*

Results

Excluded Participants

Three participants (out of 56) expressed suspicion at debriefing and were not included in the analyses presented below. Only 4 actors picked a White suspect as the most likely culprit. Because we did not expect participants to be threatened by the rebel if they had not picked the African American suspect, we excluded these 4 participants from subsequent analyses (leaving 49 valid participants). We repeated the main analyses with all 56 original participants and found that patterns and significance level were the same with all participants.

Is the Rebellion Moral?

We first compared ratings of rebel and obedient others by observers (who, being more disinterested, should give us more of a baseline) on the *immoral–moral* semantic differential. Observers thought that rebels were significantly more moral ($M = 1.17$, $SD = 1.19$) than obedient others ($M = 0.31$, $SD = 0.85$), $t(23) = 2.08$, $p < .05$. This suggests that rebellion in the police decision task had a moral flavor for participants. Interestingly, looking at morality ratings for actors, there was absolutely no difference between rebels ($M = 0.33$, $SD = 1.30$) and compliant others ($M = 0.33$, $SD = 0.39$), as if actors refused to give rebels moral credit for their good deed.

Attraction

We created an index of attraction to the target by averaging scores on liking as a friend, as a roommate, on a project, and respect (Cronbach's $\alpha = .82$) and conducted a Rebellion (rebel, obedient) × Role (actor, observer) ANOVA on this index. As predicted, we found a significant interaction between role and rebellion, $F(1, 45) = 4.38$, $p = .04$, $MSE = 2.88$, partial $\eta^2 = .09$, due to the fact that actors marginally preferred the obedient other ($M = 0.50$, $SD = 1.34$) to the rebel ($M = -0.67$, $SD = 2.03$), $t(45) = 1.68$, $p < .10$, $d = 0.71$, whereas observers did not have a significant preference between the rebel ($M = 0.98$, $SD = 1.64$) and the obedient target ($M = 0.12$, $SD = 1.71$), $t(45) = 1.27$, $p = .21$, $d = -0.53$. Neither of the main effects was significant (both $ps > .20$, partial $\eta^2 < .04$).

Open-Ended Responses

Although we have focused on continuous variables above, nowhere is the effect better grasped than by simply looking at the

trait that participants suggested to describe the rebel: Whereas observers called the rebel "strong" (twice), "strong-minded," "independent," "decisive," "fair-minded," "socially conscious," "adamant," or "not racist" (with only three observers calling the rebel "proud," "blunt," or "self-righteous"), actors viewing *the exact same stimulus* called the rebel "self-righteous" (twice), "defensive," "opinionated," "confused," "easily offended," and "racist" (with only four actors calling the rebel "stalwart," "quirky," "bold," or "intelligent"). In fact, valence ratings of these traits in Studies 1 and 2 by judges blind to condition yield the same significant cross-over interaction. (We do not present these analyses only because they are entirely redundant with the Likert scale data.)

Agency and Communion

As before, we conducted a principal axis factor analysis with a Promax rotation on the trait ratings. The scree plot this time suggested three factors explaining 54% of the variance, and using the criterion that a trait needed to load on one (and only one) factor .45 or higher to be included, we formed a first factor (agency) by averaging *confident*, *high self-esteem*, *independent*, *strong*, *moral*, *honest*, *fair*, and *active* ($\alpha = .88$), a second factor (social skill) by averaging *polite*, *pleasant*, *intelligent*, and *mature* ($\alpha = .80$), and a third factor (communion) by averaging *warm*, *nice*, and *generous* ($\alpha = .72$). In this new set of traits, *kind* did not join any factor, loading lower than .34 on all three. Agency correlated with social skills, $r(47) = .36$, $p = .01$, but not significantly with communion, $r(47) = .17$, $p = .25$, whereas communion and social skills correlated highest, $r(47) = .40$, $p = .004$.

The Role × Rebellion ANOVA conducted on each of these three factors revealed a significant interaction on agency, $F(1, 45) = 4.01$, $p = .05$, $MSE = .95$, partial $\eta^2 = .08$, and a marginal one on communion, $F(1, 45) = 3.38$, $p = .07$, $MSE = .51$, partial $\eta^2 = .07$, but none for social skills, $F(1, 45) < 1$, ns, $MSE = 1.06$, partial $\eta^2 = .02$. None of the main effects were significant (all ps >.15, all partial η^2s <.05). Tests of simple effects suggest that the agency interaction comes from observers rating the rebel more agentic than the compliant other, $t(45) = 2.24$, $p = .03$, whereas actors did not, $t(45) = 0.61$, $p = .55$, and that the marginal interaction on communion comes from observers perceiving the rebel as nicer than the compliant other, $t(45) = 2.27$, $p = .03$, whereas actors did not, $t(45) = 0.38$, ns.

Discussion

Study 2 replicates the cross-over pattern observed in Study 1: Whether participants liked someone taking a stand against a racist situation more than a compliant target depended on participants' own involvement in the task (supporting Hypothesis 1). Actors who had already gone along with the racist task liked the rebel less than an obedient other (marginally supporting Prediction 1a), whereas observers who had not participated in the task yet seemed to prefer the rebel, but this difference was not significant here (Prediction 1b was not supported).

In Study 2, the significant interaction on the attraction variable therefore seems to be driven by actors liking and respecting the rebel (marginally) less, whereas in Study 1, not only did actors like the rebel less than the compliant other, but observers also liked the

rebel significantly more. It is of interest that, although observers in Study 2 did not report liking and respecting the rebel any more, they did rate him or her significantly higher on the agency and the communion composites, driving a significant interaction in both cases. They also thought that the rebel was more moral than an obedient other. Because the placement of the trait ratings and the measures of attraction were reversed in Study 2, it is possible that for observers to start appreciating rebels, they need to reflect on the implication of their behavior, as they do when they rate their personality. Seeing the attraction measures first in Study 2, they did not report attraction for the rebel, though they did attribute more positive qualities to him or her once they got to the trait rating task.

One issue that was not addressed in Study 2 was the identity of the target. In particular, participants were not provided any information about the target's gender and ethnicity, and this could have led participants to make different assumptions on the basis of their condition. These demographic assumptions might mediate the effect. If, for example, participants assumed that the rebel standing against racism was African American, then it would give a different meaning to their act and change the nature of the threat for non-African American respondents. The fact that our finding is an interaction (i.e., that the rebel is sometimes liked more, and sometimes liked less) reduces this concern, but in the next studies, we made sure to clearly identify the speaker's gender and ethnicity in order to eliminate this issue altogether. In Study 3, we specified that the target was a man, and we used only male participants, and in Study 4, we used men and women but always matched gender of target with that of the participant.

Having demonstrated the moderating role of involvement on reactions to moral rebels in two studies, in the final two studies we endeavored to demonstrate the psychological processes involved in bringing about resentment. Accordingly, Study 3 is a mediation study in which we predict that the rejection of moral rebels can be explained by the fear of being rejected by them (Hypothesis 2), whereas Study 4 is a moderation study that demonstrates the role of self-threat (Hypothesis 3).

Study 3: The Mediating Role of Imagined Rejection

The goal of Study 3 was to show that the rejection of moral rebels is a reaction to people's perception that the rebels would reject them, presumably because they see them as less moral (Hypothesis 2). Though such judgment can be aversive in other domains, the particular sting of moral reproach (Sabini & Silver, 1982), as well as the centrality of morality in many people's self-concept (Park et al., 2007), makes fear of implicit moral reproach a likely trigger of resentment. To test whether the sting of moral rebels could be explained by the fear of disapproval, we tested the mediating role of imagined liking and respect in the effects described so far. We predicted that actors would expect to be rejected by rebels (Prediction 2a) and that resentment results from this rejection threat, and therefore that fear of rejection would mediate the rejection of moral rebels by actors (Prediction 2b).

Instead of asking participants directly about imagined moral reproach, which might be difficult for participants to articulate, or even self-threat, which we assumed participants would be reluctant to admit if probed head-on, we tapped into imagined rejection by turning around the liking and respect questions used in Studies 1

and 2 and asking participants who had gone along with the task (actors) how they thought they would be seen by the person whose completed the questionnaire they rated.

To assess other possible factors leading to the rejection of moral rebels, we also asked participants who had done the task first (actors) to answer several other ancillary questions. First, we asked participants how satisfied they were with their choice right after making it, and then again after seeing the other's choice. Second, we included several items that addressed how much rebels invalidated excuses by reducing the perceived pull of situational demands.

Finally, to make sure that the effect observed in Studies 1 and 2 did not result from gender stereotypes (e.g., if the rebel is more likely to be seen as a man), or cross-gender perception (e.g., if men and women are differentially attracted by an agentic male), in Study 3, we used only male participants and always specified that the target other was a man as well.

Method

Participants

One hundred thirty-two male undergraduates at a large U.S. private university (60 Whites, 53 Asians, 8 non-Black multiracial, 7 Hispanics, 1 Native American, and 3 unreported) were recruited by a White female or an Hispanic male experimenter to fill out a survey at various campus locations.

Procedure

The design and procedure for Study 3 was the same as Study 2, except for four main changes. First, to save time and reduce suspicion, participants in the observer condition were not asked to fill out the police decision task themselves. These responses were of little use to the analysis in Study 2, and the repetition of the questionnaire raised suspicion in participants who assumed that we were trying to influence their own response by showing them someone else's first. Second, we added a brief demographic survey at the top of the police decision task response sheet, enabling us to inform participants that the person who allegedly completed the form was a White male. Third, as mentioned above, we used only male participants. Fourth, and most important, participants in the actor condition completed an extra sheet of process questions after they had rated the other participant.

Materials

The materials used in Study 3 were the same as in Study 2, except for a few changes. As mentioned above, we added basic demographics (gender, race, and age) at the top of the police decision task—the target was now identified as male, White, and age 19. In the liking questions, we substituted the roommate question used in Studies 1 and 2 with the question "How much would you like to have a conversation with the other participant?" To shorten the procedure, we also removed all trait ratings, though we left the open-ended question. We added at the bottom of this first sheet the statement "I am very happy with my choice of a suspect," followed by a 7-point scale ranging from −3 (*strongly disagree*) to +3 (*strongly agree*). Actors received an additional sheet after seeing the choice of the other, where first they used the

same agreement scale as above to rate the following statements: "I am very happy with my choice of a suspect" (a second time), "My choice is representative of my attitudes and values," "Anyone would have picked the same suspect as I did," "The information presented was sufficient to lead me to the correct choice," and "I had no other choice but to select the most obvious suspect." Finally, they answered all three liking questions again, as well as the respect question, but from the point of view of the other participant (e.g., "After seeing your answers on the police decision making task, how much would the other participant like you as a friend?"), using the same 11-point scales used to rate the other.

Results

Excluded Participants

Only 2 participants (out of 132) expressed suspicion. They were excluded from the rest of our analyses. Thirteen actors picked one of the White targets as the likely suspect. Because our hypothesis was predicated on the fact that actors have engaged in the problematic behavior, we did not predict resentment for individuals having in some way already taken a stand. For that reason, we excluded from subsequent analyses these 13 actors, leaving 117 valid participants, but as in previous studies, we repeated the main analyses with all 132 initial participants, and none of the patterns or significance levels changed with all participants included.

Attraction

As in previous studies, we computed an aggregate variable of attraction by averaging the three liking variables and the respect variable (Cronbach's α = .86), with low values indicating more rejection. We conducted the Role × Rebellion ANOVA on this aggregate variable and found a marginal main effect of role, $F(1, 113) = 3.24$, $p = .08$, $MSE = 3.12$, qualified by the predicted significant two-way interaction, $F(1, 113) = 5.58$, $p = .02$, partial $\eta^2 = .05$. Overall, actors preferred obedient others ($M = 1.63$, $SD = 1.15$) to rebels ($M = 0.53$, $SD = 2.27$), $t(113) = 2.33$, $p = .02$, $d = 0.61$, whereas observers showed little preference ($M = 0.27$, $SD = 1.47$ vs. $M = 0.71$, $SD = 1.923$), $t(113) = .98$, ns, $d = −0.27$.

Imagined Attraction

We next turn our attention to the imagined attraction measures collected for actors (2 actors left these blank and are not included in these analyses, leaving 54 valid). We created an aggregate score of *imagined attraction* by averaging the three scores of imagined liking and imagined respect by the other, all ranging from −3 to +3 (Cronbach's α = .88), with low scores indicating more imagined rejection. As predicted (Prediction 2a), actors expected to be rejected by a rebel ($M = −0.88$, $SD = 1.94$), but not by an obedient other ($M = 1.5$, $SD = 1.13$), $t(52) = 5.33$, $p < .001$.

Mediation Analyses

We tested whether imagined attraction mediated the impact of rebellion on attraction for actors (rebellion in the subsequent regression analyses is coded such that 0 is obedient and 1 is rebel), as illustrated in Figure 2. Actors rejected rebels more than obedient
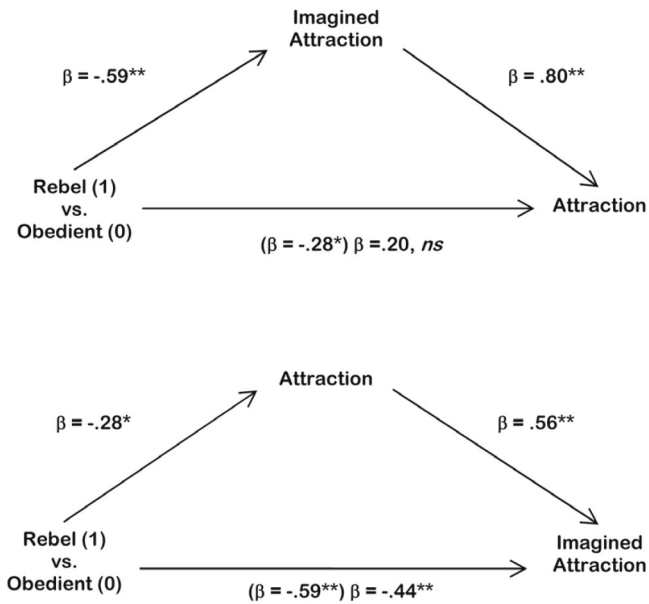
Figure 2. Mediation analyses for actors in Study 3. Coefficients in parentheses correspond to the initial effect to be explained. The top panel shows that imagined attraction mediates the effect of rebellion on attraction; the bottom panel tests the reverse causal path and shows that attraction does not mediate the impact of rebellion on imagined attraction. $^*p < .05.$ $^{**}p < .01.$

others ($\beta = -.28$), $t(52) = -2.12$, $p = .04$, and thought that rebels would reject them more than would obedient others ($\beta = -.59$), $t(52) = -5.33$, $p < .001$. When we used both the rebellion manipulation and imagined attraction to predict attraction, imagined attraction was a significant predictor ($\beta = .80$), $t(51) = 6.50$, $p < .001$, whereas the rebellion manipulation was no longer significant ($\beta = .20$), $t(51) = 1.58$, *ns*, and this reduction was significant in a Sobel test ($z = 4.12$, $p < .001$).

Furthermore, to make sure that imagined attraction was not actually a consequence of attraction (especially because it was measured later), we also tested an alternative mediational model in which attraction mediated the relationship between rebellion and imagined attraction (see Figure 2, bottom panel). We already knew that rebellion significantly affected both imagined attraction and attraction (see above). We added a regression predicting imagined attraction with rebellion and attraction and found that attraction was indeed a significant predictor ($\beta = .56$), $t(51) = 6.50$, $p < .001$, but also that the rebellion manipulation was still a highly significant predictor ($\beta = -.44$), $t(51) = -5.01$, $p < .001$. This analysis supports our confidence that imagined attraction is the mediator of attraction, and not the other way around.

Happiness With Choice and Situation Blame

Actors did not express less happiness with their own choice after seeing a rebel ($M = 0.53$, $SD = 1.43$) than after seeing an obedient other ($M = 0.67$, $SD = 1.40$), $t(52) = 0.34$, *ns*, even when looked at as a difference from the measure taken just after their choice, $t(52) = 0.37$, *ns*.

Next, we computed a score of situation blame by averaging the three choice items ("Anyone would have picked the same suspect as I did," "The information presented was sufficient to lead me to the correct choice," and "I had no other choice but to select the most obvious suspect") into an aggregate that could range from -3 to $+3$ (Cronbach's $\alpha = .61$). Actors who had seen someone rebel were slightly less likely to blame the situation for their own choice ($M = 0.24$, $SD = 1.22$) than if they had seen someone obey ($M = 0.76$, $SD = 1.29$), but this difference was not significant, $t(52) = 1.52$, $p = .14$, $d = 0.42$.

Discussion

As in Studies 1 and 2, Hypothesis 1 was supported in Study 3: Reactions to moral rebels depended on participants' own involvement in the task. Actors who had gone along with the task liked a rebel less than a compliant other (supporting Prediction 1a), whereas mere observers did not show this preference (Prediction 1b was, as in Study 2, not supported). By tapping into imagined rejection, Study 3 enabled us to go beyond demonstrating the rejection of moral rebels to start looking at the causes leading up to it. The mediation analyses suggest that the rejection of rebels by actors can be explained as a reaction to the sense that one would be rejected by the person taking the moral high ground, as demonstrated by the facts that actors did expect to be liked less by the rebel (Prediction 2a), and that the effect of rebellion on rejection was mediated by imagined rejection (Prediction 2b). Furthermore, a test of the alternative model suggests that the reverse is not true: It is unlikely that imagined rejection is the result of rejection, as the impact of rebellion on imagined rejection is still highly significant when rejection is included in the analysis.

Imagined rejection thus seems to be an important factor in the rejection of rebels. We found little support, however, for the idea that rebels make individuals less happy with their own decision or that rebels make individuals more aware that they could have gone beyond the pressures of the situation, although it is possible that these variables are less accessible for self-report, less willingly revealed, or were measured too late in the procedure. Going back to the imagined rejection mediator, it is striking that, as we pointed out earlier, participants had no reason to believe that they would ever meet the rebel or, for that matter, that the rebel would ever see their response and thus get a chance to reject them. Thus, the mediating cognition is not what the rebel *will think*, but what the rebel *would have thought* were he still around. The fact that the actual presence of the rejecting rebel is not necessary to yield rejection suggests that imagined rejection may be less a threat to an actual social relationship than a threat to one's personal sense of integrity and self-worth. We propose that the thought that someone would reject you constitutes a threat to one's "moral and adaptive adequacy" (Steele, 1988)—and that this threat triggers rejection. Study 4 tests this piece of the model.

Study 4: Buttressing the Self

Study 3 suggests that moral rebels are resented by actors because of the perception that they would, if they could, reject others who do not take a similar position. This fits our model, which posits that the imagined rejection perceived in moral rebels is a threat to one's sense of adequacy (Hypothesis 3). As we have

noted, the moral rebel in our studies is actually not in a position to judge participants, but we argue that the simple fact that someone could look down on them can shake participants' self-confidence, and it is this sting that triggers the rejection. Having shown the role of imagined reproach in Study 3, in Study 4 we proceeded to show the implication of the self. Specifically, we reasoned that if the rejection of moral rebels indeed results from shaking participants' self-confidence, then actors should not feel a need to reject rebels if they have been secured in their sense of being a good, effective person (Prediction 3a).

Spencer et al. (2001) showed that after receiving bogus negative feedback on an intelligence test, participants chose to listen to a poorly performing peer in preparation for a later interview. However, when participants had had the chance to write about an important value, they preferred to listen to a highly performing peer. Apparently, this opportunity to self-affirm made an impressive peer more palatable, despite the recent setback of the intelligence test. Along similar lines, and in line with self-affirmation theory (Sherman & Cohen, 2006; Steele, 1988), we predicted that buttressing participants' sense of adequacy by giving them a chance to reflect on one of their important values or qualities would shield them from the sting of imagined rejection and would therefore reduce the need to put down moral rebels (Prediction 3a).

In essence, we predicted that self-affirmed actors in Study 4 would resemble uninvolved observers in Studies 1–3. Prior research suggests that self-affirmation can reduce the tendency to deny or skew information that is threatening to one's beliefs, and thus reduce the impact of one's initial point of view on the processing of new information (e.g., Cohen, Aronson, & Steele, 2000). We predicted that self-affirmation would similarly give actors some distance, both reducing their need to reject the rebel (Prediction 3a, above) and allowing them to recognize the value of the rebel's behavior as a moral, agentic choice—possibly to the point of questioning their own (Prediction 3b).

A secondary goal of Study 4 was to determine whether the rejection of a moral rebel serves to reduce the kind of negative affect or psychological discomfort assumed to accompany cognitive dissonance (Festinger, 1957). We thus added measures of self-reported affect and discomfort, varying whether they were presented before or after the opportunity to put down the rebel. The placement of these measures could be important, because if a process serves to reduce discomfort, then discomfort should be high when measured before it but lower when measured after (cf. Elliot & Devine, 1994).

Method

Participants

Seventy-nine undergraduates at a large U.S. private university (47 Whites, 17 Asians, 9 Hispanics, 3 Native Americans, 2 non-Black multiracial, and 1 Iranian; 52 men, 27 women) came to the laboratory to complete the research participation requirement of an introductory psychology class or in exchange of a payment of $8. The experimenter was a White man.

Design

Participants were randomly assigned to one of three conditions: obedient, rebel control, or rebel self-affirmation. All participants in this study were actors (i.e., completed the police decision task first) and saw a compliant other in the first condition, or a rebel other in the other two. Before rating the other, participants wrote a personal essay in the self-affirmation condition or listed foods in the rebel control and obedient conditions. We also included two different orders within the rebel control condition to look at a possible order effect on affect measures (cf. Elliot & Devine, 1994), but we collapsed these two conditions into one for simplicity after discovering that this order made no difference (see the *Placement of affect measures* section).

Procedure

Participants came to the laboratory to take part in a decision-making study and were told that they would make decisions, form impressions of other students, and fill out some mood and emotion questionnaires. They first completed the police decision task from Study 3. The experimenter, sitting with his back toward participants for the duration of the study, then came by with a clipboard and ostensibly made a note of the participant's choice in the task. After this, participants wrote a personal essay or listed foods they had eaten (see below). After 8 min, the experimenter told participants that they would switch to forming impressions of a past participant and gave them a blank rating sheet and a completed version of the police decision task (containing the rebel vs. obedient manipulation) ostensibly taken randomly from a stack of completed questionnaires. The other participant was always White, age 19, and matched in gender with the participants. Participants' own completed decision task was left on the table, face up, next to the one allegedly completed by the other student—and it was left there until the end of the study to make sure it remained salient to participants. After indicating their impression of the other participant, participants rated their affective state on a list of traits (half the participants in the rebel control condition did this first) and answered the questions about the situation and about expected rejection introduced in Study 3.

Self-Affirmation Manipulation

In the rebel self-affirmation condition, participants read the following instructions:

> Please write about a recent experience in which you demonstrated a quality or value that is very important to you and which made you feel good about yourself. Examples of "personally important values or quality" might include (but are not limited to) artistic skills, sense of humor, social skills, spontaneity, athletic ability, musical talent, physical attractiveness, creativity, business skills, or romantic values.

Participants named their chosen quality or value, rated its personal importance on a scale from 0 (*not important at all*) to 4 (*extremely important*), then described their experience on the blank lines provided on the rest of the page in the remaining 8 min. Participants in the obedient or rebel control conditions instead read the following instructions: "Please describe everything you have eaten or drunk in the past 48 hours. Do not worry about things you find

Table 1
Condition Means, Standard Deviations, and Omnibus F Tests for the Main Variables in Study 4

Variable	Obedient		Rebel control		Rebel self-affirmation		Omnibus test	
	M	SD	M	SD	M	SD	$F(2, 64)$	MSE
Attraction	1.88$_a$	1.38	0.02$_b$	2.38	2.54$_a$	1.95	10.17**	4.08
Morality	1.11$_a$	1.66	1.41$_a$	1.30	2.42$_b$	1.02	5.12**	1.80
Communion	0.83$_a$	0.89	0.25$_b$	0.98	0.87$_a$	0.62	3.99*	0.74
Agency	1.34$_a$	0.87	1.44$_a$	1.09	2.10$_b$	1.02	3.31*	1.02
Imagined attraction	1.00$_a$	2.03	−1.52$_b$	1.38	−0.58$_b$	2.00	11.68**	3.12
Happiness with choice	1.32$_a$	1.67	1.24$_a$	1.41	0.26$_b$	1.66	2.87†	2.42
Situation blame	1.42$_a$	1.19	1.03$_a$	1.36	−0.07$_b$	1.62	5.98**	1.94
n	19		29		19			

Note. Means that do not share the same subscript differ at $p < .05$, using t tests with 64 degrees of freedom.
$^*p < .05$. $^{**}p < .01$. $^\dagger p < .10$.

yourself unable to remember," followed by blank lines on the rest of the page.[3]

Measures

Ratings were similar to previous studies, with some small changes. Participants indicated how much they would like to work on a project with the other participant and how much they would like him or her as a friend, both on an 11-point scale ranging from −5 (*dislike very much*) to +5 (*like very much*), with an unlabeled midpoint of 0. They indicated how much they respected the other participant on a similar 11-point scale ranging from −5 (*despise a great deal*) to +5 (*respect a great deal*), then indicated the personality trait that they thought best described the other participant (open ended), and finally rated the other participant on the 14 semantic differentials on 7-point scales with a midpoint of 0 (*neither*) and with the following poles: *stupid–intelligent, strong–weak, unpleasant–pleasant, funny–not funny, confident–insecure, immature–mature, active–passive, cruel–kind, nice–awful, cold–warm, honest–dishonest, unfair–fair, moral–immoral,* and *dependent–independent.*

On the following page, participants rated their own morality relative to other students on campus on a scale ranging from 0% (*you are the least moral student on campus*) to 100% (*you are the most moral student on campus*) and a midpoint of 50% (*average*). Then, under the heading "mood and emotion survey," they answered the question "How do you feel right now?" by rating each of the following 24 states on a scale ranging from 1 (*does not apply at all*) to 7 (*applies very much*): disappointed with myself, good, uncomfortable, happy with myself, determined, annoyed with myself, happy, guilty, comfortable with myself, peaceful, uneasy, disgusted with myself, excited, pleased with myself, energetic, angry with myself, bothered, friendly, dissatisfied with myself, fatigued, self-critical, optimistic, secure with myself, and lonely.

Finally, participants rated his or her agreement with the following items introduced in Study 3 on a 7-point scale ranging from −3 (*strongly disagree*) to +3 (*strongly agree*): "I am very happy with my choice of a suspect"; "My choice is representative of my attitudes and values"; "Anyone would have picked the same suspect as I did"; "The information presented was sufficient to lead me to the correct choice"; and "I had no other choice but to select

the most obvious subject." On the same page, they indicated how the other participant would like and respect them after seeing their own answers on the police decision making task, using the first three scales used to rate the other participant.

Results

Table 1 displays means and standard deviations of the main variables, analyzed below, as well as one-way ANOVAs, mean square errors, and pairwise comparisons between means.

Preliminary Analyses

Suspicion. Eleven participants expressed suspicion during debriefing and were excluded from the following analyses. Only 1 participant was excluded for not picking the African American suspect (Steven) in the police decision task, leaving 67 valid participants (44 men, 23 women). As before, we repeated the main analyses with all 79 participants included, but patterns and significance levels were the same with all participants included.

Placement of affect measures. The placement of affect measures in the rebel control condition did not have a significant impact on any of the variables analyzed below. In particular, order did not significantly affect the attraction composite, $t(63) = 0.48$, *ns*. For ease of presentation, we collapsed the two orders into the rebel control condition in subsequent analyses and only mention order again when it is most relevant, that is, in the analysis of affect measures.

Self-affirmation manipulation checks. Participants in the rebel self-affirmation condition confirmed that their chosen value was very important ($M = 3.42$, $SD = 0.61$) on a 5-point scale ranging from 0 (*not important at all*) to 4 (*extremely important*). The values chosen varied from physical (fitness, athletic ability [4 times]) and other skills (performances as a poet, artistic skills-creativity, business skills, social skills) to desirable personality

---

[3] This manipulation was adapted from Cohen et al. (2000, Study 1), who noted that they "chose this control condition (instead of one that asked participants to write about an unimportant value) because students tend to turn almost any self-reflective task into a self-affirming one" (p. 1154).

traits (patience, spontaneity, romantic values, keeping contact with friends and family), with only a minority of morality-related traits (loyalty [twice], honesty, generosity, ability to help others, and concern/care for someone in need). We counted the number of words used by participants and found that participants did not use significantly more words in the rebel self-affirmation condition ($M = 115$, $SD = 39$) than in the rebel control condition ($M = 110$, $SD = 51$), $t(64) = 0.69$, *ns*. By this rough metric at least, we had no reason to believe that participants in the self-affirmation condition expended more effort than participants in the rebel control condition.

Perception of the Other

Attraction (Table 1). The three groups differed significantly on an attraction composite of liking to work on a project, liking as a friend, and respect ($\alpha = .88$). Replicating previous studies, nonaffirmed actors liked the rebel significantly less ($M = .02$) than the obedient other ($M = 1.88$; see Figure 3). More important, self-affirmed participants showed no such rejection of the rebel ($M = 2.54$).[4]

Morality (Table 1). Nonaffirmed actors did not rate a rebel significantly more moral ($M = 1.41$) than an obedient target ($M = 1.11$), but self-affirmed participants did ($M = 2.42$).

Agency and communion (Table 1). The scree plot in a principal axis factor analysis with a Promax rotation on the 14 bipolar trait items suggested two factors, together accounting for 45% of the variance. A criterion loading of .40 excluded "funny" (both loadings below .16) but otherwise yielded a nonoverlapping solution: *communion* ($\alpha = .84$) comprising nice, kind, pleasant, warm, fair, and mature; and *agency* ($\alpha = .82$, correlated .46 with communion, $p < .001$) comprising strong, independent, intelligent, honest, secure, moral, and active. As is apparent in Table 1, communion was significantly lower in the rebel control condition ($M = 0.25$) than in both the obedient ($M = 0.83$) and the rebel self-affirmation condition ($M = 0.87$). In contrast, agency was significantly higher in the rebel self-affirmation condition ($M =$

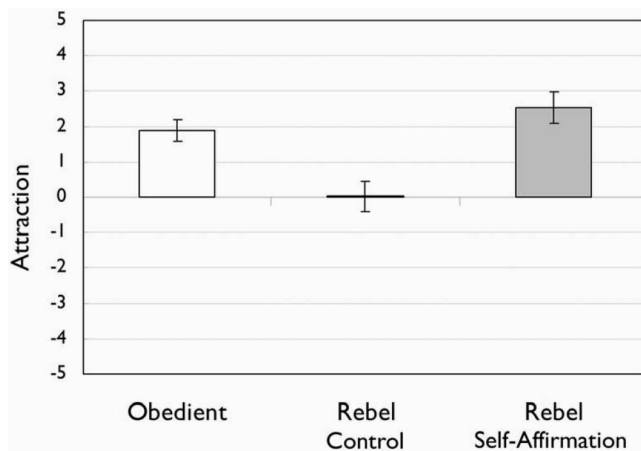*Figure 3.* Attraction means (± 1 *SE*) per condition in Study 4. Whereas a rebel other is still liked less than an obedient other in the first two conditions, self-affirmation increases liking for the rebel in the third condition.

2.10) than in both the compliant ($M = 1.34$) and the rebel control condition ($M = 1.44$).

Imagined Attraction

Mediation of rebel rejection. We created a composite for imagined attraction by averaging the three questions asking the participant to imagine how much the other person would like and respect him or her ($\alpha = .87$) and tested first whether it mediated the difference between the obedient and the rebel control conditions. Using a dummy code pitting the rebel control condition (1) against the obedient condition (0), rebels were expected to like participant less ($\beta = -.60$), $t(46) = -5.12$, $p < .001$; and when both rebellion and imagined attraction were entered to predict attraction, imagined attraction was a significant predictor ($\beta = .38$), $t(45) = 2.16$, $p = .04$, whereas the rebellion manipulation, which had been a significant predictor on its own ($\beta = -.41$), $t(46) = -3.07$, $p = .004$, was no longer so with the mediator included ($\beta = -.20$), $t(45) = -1.24$, $p = .22$, Sobel $z = -2.00$, $p < .05$. As in Study 3, the reverse causal order was not supported because rebellion still predicted imagined attraction when attraction was controlled for ($\beta = -.49$), $t(45) = -3.96$, $p < .001$.

Self-affirmation and imagined attraction. Participants in the rebel self-affirmation condition expected a lukewarm reaction from the other participant ($M = -0.58$), definitely cooler than in the obedient condition ($M = 1.00$), $t(64) = 2.76$, $p = .01$, and only marginally warmer than in the rebel control condition ($M = -1.52$), $t(64) = 1.80$, $p = .08$. Does this marginal boost in imagined attraction explain why self-affirmed participants did not reject a rebel? To test this, we conducted a mediation analysis using a dummy code pitting the rebel control condition (0) against the rebel self-affirmation condition (1). As above, affirmation marginally increased imagined attraction ($\beta = .27$), $t(46) = 1.93$, $p = .06$, and when both variables were entered in the equation, imagined attraction was significant ($\beta = .36$), $t(45) = 2.95$, $p = .005$, but self-affirmation was still a highly significant predictor ($\beta = .39$), $t(45) = 3.19$, $p = .003$, and not significantly reduced by the inclusion of imagined attraction (Sobel $z = 1.61$, $p = .11$). Thus, the more clement view of rebels by self-affirmed participants does not seem to result from expecting to be liked better.

Affect

Identifying factors. A principal axis factor analysis with a Promax rotation on the 24 affect items suggested three factors (scree plot method) capturing 56% of the variance. A loading cutpoint of .45 excluded "friendly" (loadings $< .42$) and "uncomfortable" (loadings $< .34$), but otherwise yielded nonoverlapping factors: *negative affect* ($\alpha = .90$), which comprised disgusted with myself, angry with myself, dissatisfied with myself, disappointed

---

[4] Note that participants who self-affirmed with a moral value ($n = 6$) were less positive about the rebel ($M = 1.61$, $SD = 2.02$) than participants who self-affirmed with a nonmoral value ($n = 13$; $M = 2.97$, $SD = 1.83$, $d = 0.76$), though this difference was not significant with so few participants, $t(17) = 1.46$, $p = .16$. This trend is consistent with previous research suggesting that self-affirmation using the domain that will later be challenged is less effective because it highlights inconsistency (Aronson, Blanton, & Cooper, 1995).

with myself, annoyed with myself, fatigued, guilty, bothered, lonely, and self-critical; *positive affect—high arousal* ($\alpha = .89$), $r(65) = -.41$, $p = .001$, with negative affect, comprising energetic, excited, happy, pleased with myself, determined, good, happy with myself, and optimistic; and *positive affect—low arousal* ($\alpha = .82$), $r(65) = -.65$, with negative affect; and, $r(65) = .59$, with positive affect—high arousal (both $p < .001$), comprising secure with myself, peaceful, comfortable with myself, and uneasy (reversed scored).

No mean differences between groups. As mentioned in the preliminary analyses, the placement of these measures in the rebel control condition did not affect any of these measures significantly, all three $ts(63) < .5$, $ps > .6$. After collapsing these two orders into the rebel control condition, we did not find any significant difference between the three groups either, all three $Fs(2, 64) < 2.0$, $ps > .15$. In particular, the rebel self-affirmation condition did not generate more positive affect than other conditions, ruling out the possibility that it was merely a mood manipulation (Sherman & Cohen, 2006). Given the high correlations between measures, we created an overall affect composite averaging all 24 affect items after reversing the positive items ($\alpha = .94$), but again we observed no significant difference between group means, $F(2, 64) = 1.63$, $p = .20$, or between orders within the rebel control condition, $t(63) < .2$, *ns*.

How does it feel to reject the other? To test whether different feelings might be associated with attraction in the three conditions, we conducted a regression predicting attraction with the overall affect composite (standardized), two dummy codes capturing the rebel control and the rebel self-affirmation condition, and two product terms of the dummy codes and the standardized affect composite. Reflecting that rebels are liked less than obedient others, the dummy code for the rebel control condition was a significant predictor, $t(61) = -2.93$, $p = .005$. More germane to the matter at hand, the product between this dummy code and the affect composite was also a significant predictor, $t(61) = 2.80$, $p = .007$ (all other $ts < 1.3$, *ns*): Whereas rejecting an obedient other was associated with negative affect, $r(17) = -.45$, $p = .05$, rejecting a rebel was associated with positive affect, $r(27) = .41$[5], $p = .03$. For self-affirmed participants, his or her reaction to the rebel seems detached from affect, $r(17) = .04$, *ns*, though the fact that the corresponding interaction term was not significant in the regression makes unclear whether this apparent departure from the obedient condition is reliable.

Secondary Measures

Ratings of one's moral standing. When asked for an estimate of their morality percentile, participants reported on average that they were more moral than two thirds of their peers ($M = 67\%$, $SD = 15\%$), but this did not vary by condition, $F(2, 64) < 1$, *ns*.

Happiness with choice (Table 1). As in Study 3, participants did not report less happiness with their choice of suspect in the rebel control condition ($M = 1.24$) than in the obedient condition ($M = 1.32$), but in line with their tendency to face the music, self-affirmed participants did ($M = 0.26$).

Situational blame (Table 1). Similarly, the tendency to blame the situation (computed by averaging "I had no other choice but to select the most obvious suspect"; "Anyone would have picked the same suspect as I did"; and "The information presented was

sufficient to lead me to the correct choice" [$\alpha = 69$]) was significantly lower in the rebel self-affirmation condition ($M = -0.07$) than in the rebel control condition ($M = 1.03$), itself similar to the obedient condition ($M = 1.42$).

Discussion

Study 4 was designed to demonstrate that the rejection of moral rebels is a reaction to a threat to the self. Whereas we once more observed rebel rejection among actors (supporting Prediction 1a) in the two cells in which judging the other was preceded by a mundane task (listing what one had eaten for 48 hr prior to the study), rebels were no longer rejected if participants first had a chance to feel secured in their sense of moral and adaptive adequacy by recalling a recent experience when they demonstrated an important quality or value (supporting Prediction 3a). Self-affirmed actors liked and respected a moral rebel as much as actors liked and respected a compliant other, looking very much like uninvolved observers in Studies 1–3 who were not threatened by the rebel's stance.

The fact that secure participants did not feel a need to put down moral rebels suggests that when actors do put down moral rebels, it results from a threat to one's sense of self-worth and integrity. As in Study 3, we found that the rejection of moral rebels was mediated by imagined rejection (supporting Hypothesis 2); self-affirmed participants expected marginally less rejection, but that did not explain the difference. They knew that rebels would probably not hold them in high regard, yet they did not seem to care—they still did not mind them. In fact, they were the only group for whom attraction/rejection was not related to how they were feeling.

Self-affirmed participants' ability to take stock of the rebel's stance without lashing back or denying its value went beyond merely not minding rebels despite imagined rejection. In support of Prediction 3b, self-affirmed participants seemed better able to give credit when credit was due: In contrast to participants in the rebel control condition, they saw the rebel as particularly moral and agentic, reported being less happy with their choice than participants seeing an obedient other, and even saw that they might not have been as constrained by the situation as they thought at the time.

Does this clemency and clear-sightedness in the self-affirmation condition result from mere distraction? Did it make participants lose sight of the rebel they were rating, did it make them forget about their own choice, or did it engage them more than the control condition? Our data provide little support for either of these three possibilities. First, the essay was always written before participants saw both the completed questionnaire and the ratings sheet, so it would not have the ability to distract them from the completed questionnaire that they based their ratings on. Second, after recording their choice, we intentionally always left the completed sheet on the table facing the participants, so while doing the ratings they had three sheets in front of them: their completed questionnaire and the target's side by side, and the blank rating sheet. So

---

[5] Note that the placement of the affect measures did not matter: $r(15) = .39$ when they were placed before the attraction measures, and $r(10) = .47$ when they were placed after.

it is unlikely they forgot about either their choice or that of the rebel. Third, we made sure that the writing time was the same in all three conditions, so our best available index of engagement was to count the number of words used in the various conditions—and as reported above, participants in the rebel self-affirmation condition did not write significantly more words than participants in the food-listing condition. For these three reasons, it is unlikely that the increased attraction for rebels in the personal essay condition results from distraction. Instead, we believe that the boost in self-confidence resulting from contemplating an important feature of their identity was what enabled participants in the personal essay condition to weather the threat of the moral rebel.

General Discussion

In four studies, we investigated why those doing the right thing are not always embraced by others, and we showed that reactions to moral rebels largely depend on whether their principled stance is perceived as an implicit rejection of those who went along with the problematic situation. When fictitious rebels refused to write a deceitful speech or to collaborate in a racist decision task, they were liked more than (Study 1) or as much as (Studies 2 and 3) obedient others by uninvolved observers but less than obedient others by participants who had already gone along with the morally problematic behavior (Studies 1–4).

Study 1 showed this phenomenon using a modified induced compliance paradigm, in which low-choice participants who had written a counterattitudinal speech disliked a peer refusing to write the speech, although observers liked the rebel more than an obedient participant. In Study 2, rebels who refused to make a choice in a whodunit task in which the obvious suspect was African American were disliked by participants already implicated in the task, but not by observers, who called the rebel more moral. In both studies, the exact same behavior was judged quite differently depending on the perceiver's own involvement.

Going beyond these initial two demonstration studies, the last two studies cast light on the roots of this resentment. Study 3 demonstrated that resentment is a defensive reaction to the perception that rebels are implicitly rejecting those who do not question the situation: Indeed, the rejection of rebels was mediated by the belief that the rebel would reject others who did not rebel. Finally, in Study 4, we showed that this defensive reaction results from a threat to one's self-confidence: When participants were self-affirmed prior to seeing the rebel, they claimed to like and respect the rebel as much as nonaffirmed actors liked a compliant target—and declared that the rebel was especially moral. Together, these studies consistently demonstrate that rebels can elicit resentment and rejection in individuals who failed to take such a principled stance and who experience the rebellion as a personal rejection.

In his *Theory of Justice*, philosopher John Rawls (1999) famously posited that a fair procedure to generate just principles is to elevate individuals above "specific contingencies that put men at odds" and to assume that they are behind a "veil of ignorance" as to their position in society (p. 118). Our observer participants were in many ways in Rawls' ideal position, by not knowing how they would have acted in this situation. Not being involved, they were able to appreciate the morality and strength of character of the rebel (Study 2), and even to like and respect him or her more

(Study 1). We observed the strongest observer effect (Prediction 1b) in Study 1, when observers were not even in the laboratory but were approached 1 month later outside of the experimental situation. Note that Rawls' veil of ignorance goes beyond not knowing one's position in the social structure: "Nor, again, does anyone know his conception of the good" (p. 118). In light of our studies, one sees the wisdom of this recommendation; indeed, stepping foot in the situation as an unquestioning actor seems to change one's very "conception of the good," whereas an observer behind the veil of ignorance sees a brave moral exemplar, an actor down in the trenches sees a self-righteous pest.

Summary of Hypothesis Testing

How did the data specifically support our hypotheses, as spelled out in the introduction? Hypothesis 1, the interaction pattern, was supported in every study in which it could be tested (Studies 1, 2, and 3). Prediction 1a, that actors would dislike the rebel compared with a compliant other, was supported in nearly every study (significant in Studies 1, 3, and, 4; marginal in Study 2). Prediction 1b, that observers would prefer a compliant other to the rebel, was supported in Study 1 but not in Studies 2 and 3 (and could not be tested in Study 4). As is apparent in Table 2, which summarizes effect sizes for Hypothesis 1 and Predictions 1a and 1b, the effect predicted in Prediction 1b was actually always in the right direction (in fact, the average d observed across studies was -0.51) but consistently smaller than the effect predicted in Prediction 1a (average $d = 0.86$). It is important to realize, however, that the relative support for Prediction 1a and Prediction 1b is also a function of the stimuli used: By creating a more likable rebel, we might still obtain the predicted interaction (Hypothesis 1) because observers like the rebel significantly more (Prediction 1b), whereas actors might now rate him or her as likable as the compliant other (and Prediction 1a would seem unsupported). In fact, this is what we obtain on agency, a dimension on which our constructed rebel clearly stood above the compliant other: In Studies 1 and 2, observers rated the rebel significantly more agentic (Prediction 1b), whereas actors (in Studies 1, 2, or 4) never gave rebels credit for being agentic (Prediction 1a). Similarly, observers rated the rebel more moral when they had this opportunity (Study 2), whereas actors did not.

In regard to Hypotheses 2 and 3, that the rejection of moral rebels is a reaction to a threat to the self stemming from imagined rejection, evidence for that contention comes from at least three

Table 2

Effect Sizes on Attraction per Study and Significance Levels of the Corresponding Tests

Study	Hypothesis 1: Interaction (partial η^2)	Prediction 1a: Actors (Cohen's d)	Prediction 1b: Observers (Cohen's d)
Study 1	.16**	1.19*	−0.72*
Study 2	.09*	0.71†	−0.53
Study 3	.05*	0.61*	−0.27
Study 4		0.93**	

Note. Positive numbers on Cohen's d mean a preference for the compliant other over the rebel.
* $p < .05$. ** $p < .01$. † $p < .10$.

different sources: First, when participants were mere observers (in Studies 1–3), the same rebel target did not elicit rejection and, if anything, was more appealing than a compliant other (significant in Study 1). The importance of self-involvement was the first hint that the underlying effect was based on self-threat. Second, the mediating role of imagined rejection (Predictions 2a and 2b, supported in both Studies 3 and 4) suggests that the effect is driven by the realization that the adequacy of participants' choice could be questioned. Third, the role of self-threat may have been best demonstrated when participants who at first wrote about an important quality or value did not reject the rebel target at all (Prediction 3a, supported in Study 4). Feeling secure in their sense of adequacy, these participants were apparently not threatened by the rebel's behavior, or the fact that he or she would probably dislike them—and as a result, they did not derogate. In fact, they saw rebels as moral and agentic, expressed misgivings about their own behavior in light of the rebel's stance, and blamed their own behavior less on situational factors (Prediction 3b, supported in Study 4) than did participants who were not self-affirmed.

Alternative Interpretations

Existential Freedom

The idea that individuals are more comfortable turning a blind eye to their own freedom and attributing their problematic choices to situational pressures has a venerable past in philosophy (Sartre, 1943/1956), clinical psychology (Fromm, 1941), and social psychology (Festinger, 1957). Are rebels resented because they shatter the comfort of conformity and remind actors that they were free all along? Indeed, individuals might not have thought of rebellion as a behavioral option until rebels made them realize that it was available. As one of Milgram's (1965) obedient participants recounts at debriefing, "the thought of stopping didn't enter my mind until it was put there by the other two [rebels]" (p. 132). Although this is a compelling narrative, we find little support for it in our data. In Study 1, participants did report significantly more perceived choice after seeing the rebel, but this did not correlate significantly with rejection. When we measured it again in Studies 3 and 4, we did not observe significantly less situational blame (e.g., "I had no other choice but to select the most obvious suspect") when participants saw a rebel than when they saw a compliant other. Self-affirmed participants who saw a rebel in Study 4 did blame the situation significantly less than participants seeing an obedient other, as if they were secure enough to admit the freedom that rebels so clearly demonstrated, but nonaffirmed participants did not report less situational blame. Measuring perceived choice and experienced freedom is a notoriously difficult enterprise (see Gosling, Denizeau, & Oberlé, 2006), but the data available to us at this point do not strongly support this existential interpretation, reinforcing our confidence in our imagined rejection/self-threat model.

Feeling Less Moral Than the Rebel

Another reason why realizing the existence of a more righteous path is threatening could be that it leads individuals to question their own morality. This threat may be akin to upward social comparison (Festinger, 1954), only applied to morality. Resent-ment could be a defensive measure if perceivers feel that their own morality pales in comparison to the rebel's. Of course, individuals excel at trivializing other people's positive behavior by attributing it to social factors rather than to internal dispositions (Ybarra, 2002), so rebels will likely not be granted moral superiority in many cases; but if the moral stance is unambiguous enough, one could assume that conformists start questioning their own morality. As Nietzsche (1878/1984) wrote, "arrogance on the part of the meritorious is even more offensive to us than the arrogance of those without merit, for merit itself is offensive" (Aphorism 332)—thus sometimes it may not be the perceived self-righteousness of rebels that people are reacting to, but the threatening thought that they may be on to something. Rather than dwell on this unpleasant thought, individuals may lash out against superior others as they do when others are superior in ability (Major, Testa, & Bylsma, 1991; Salovey, 1991; Tesser, 1991), all the more strongly because of the centrality of morality in people's self-image (Allison et al., 1989; Park et al., 2007). Though this interpretation is also compelling, it does not fit the data presented here very well either. First, although unthreatened participants (observers in Study 2, affirmed actors in Study 4) did describe the rebel as more moral than the compliant other, threatened actors did not, suggesting that they were able to deny the rebel any moral credit for his or her stance. Second, when asked (Study 4), participants did not report feeling any less moral after seeing the rebel, nor did they report more self-directed negative emotions (e.g., disappointed with myself), even when this measure came before the opportunity to derogate the rebel. And finally, although a social comparison interpretation could reasonably include the presence of imagined rejection, this rejection should not mediate the effect as it does in Studies 3 and 4. Thus, we find little support for this social comparison story in our data, strengthening our faith in the self-threatening impact of imagined rejection.

Rejection by Whom?

Imagined rejection seems responsible for the self-threat that triggers rejection. But who is the source of this imagined rejection, and does it need to be restricted to the rebel? The questions used to test mediation in Studies 3 and 4 ask how participants think they would be seen by "the other participant" (i.e., the compliant or rebel other)—and as previously reported, these questions mediate the effect. It is conceivable, however, that participants would anticipate a similar rejection from a random bystander who witnessed both their obedience and the rebel's stance and that this generic imagined rejection would similarly mediate the effect. We suspect that this mediation would obtain, though it may be weaker than the observed results, and we propose that it would be compatible with our model. Our contention is that imagined rejection is threatening and that the mere existence of the rebel makes actors fear that they will be rejected or looked down upon, by the rebel or others. The imaginary and quasiprojective nature of this mediator makes it less consequential to know who is doing the imagined rejection.

Shame and Guilt

Along similar lines, recent advances in the study of shame stress the importance of imagined evaluation, even in the absence of real

observers. Tangney and Dearing (2002), for instance, wrote that "although shame doesn't necessarily involve an actual observing audience that is present to witness one's shortcomings, there is often the *imagery* of how one's defective self would appear to others" (p. 18, italics added). Tangney and Dearing also demonstrated that whereas guilt is focused on a problematic act, shame is characterized by a threat to the global self. On the basis of the self-affirmation results of Study 4, we conclude that actors reject rebels when their sense of being a good, effective person (Steele's, 1988, "moral and adaptive adequacy") is under threat. In terms of self-conscious emotions, the phenomenon of rebel rejection is therefore more related to shame than to guilt, which may explain why we observed no effect of condition on feeling "guilty" in Study 4. We might have had a difference on "ashamed" if this state descriptor had been included in the survey, though experts (e.g., Tangney & Dearing, 2002, p. 47) recognize that respondents' reluctance (or inability) to acknowledge feelings of shame, and accompanying defensive biases, make this state an elusive one to measure. To summarize, the fact that rejection is mediated by an imagery of rejection and that the global self seems under threat rather than a single act (as evidenced by the fact that rejection is eliminated when an unrelated aspect of the self is affirmed beforehand) concur to suggest that the threat triggering rejection resembles recent understandings of shame in the psychological literature.

Theoretical Contributions

The present research represents, to our knowledge, the first systematic investigation of the rejection of moral rebels. It casts light on the causes of the phenomenon, raises novel research questions, and points to directions for future investigation. It also informs a number of social psychological literatures at the same time as it draws on them:

1. It contributes to the literature on *deviance from group norm*s by showing that reactions to deviants can be largely dependent on the judge's own past behavior and by showing that the same deviant can be perceived as normative by some (seen, in fact, as more moral) and rejected by others.

2. It contributes to the literature on *social comparison* by showing that other people's behavior can be threatening not just when it appears superior, but also indirectly when it implies that others might look down on us (see also Monin, 2007).

3. It contributes to the *self-affirmation* literature by documenting a new source of self-threat and by showing a new consequence of self-affirmation, the equanimity and benevolence towards a threatening rebel that we observed in Study 4.

4. It contributes to the *attraction* literature by showing the complex interplay between exemplary behavior, imagined rejection, and self-confidence in determining who we like and who we do not.

5. It contributes to the *prejudice reduction* literature by showing that individuals who protest racist practices will have greatest appeal to those outside of the situation, unless they manage to make individuals in the situation somehow secure in their sense that they are good people.

6. Finally, it contributes to *moral psychology* by demonstrating that moral exemplars do not always receive the respect that they are supposed to inspire and by showing possible bridges between moral content and all of the mainstream social psychological

traditions touched upon in this section (see also Monin, Pizarro, & Beer, 2007).

Conclusion

We started with the puzzling observation that the same laudable rebellion was admired by some and reviled by others. The studies presented here suggest the importance of involvement as a moderator of this reaction because involved individuals perceive rebellion against a situation that they tacitly accepted as a personal rejection. Individuals reacted to this imagined rejection by rejecting rebels. Such defensiveness greatly limits the potential impact of moral leaders in society: If a minimal involvement in the situation is enough to trigger rejection, then one can imagine the consequences of a lifetime of passivity when individuals see a rebel taking a moral stance against the way others live or refusing to accept the world that others have tolerated all along. By casting light on these rejection processes, we hope to pave the way for research on the way moral rebels can be agents of change without eliciting resentment.

References

Allison, S. T., Messick, D. M., & Goethals, G. R. (1989). On being better but not smarter than others: The Muhammad Ali effect. *Social Cognition, 7,* 275–295.

Aronson, J., Blanton, H., & Cooper, J. (1995). From dissonance to disidentification: Selectivity in the self-affirmation process. *Journal of Personality and Social Psychology, 68,* 986–996.

Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs, 70*(9, Whole No. 416).

Cleaver, E. (1969). *Post-prison writings and speeches* (R. Scheer, Ed.). New York: Random House.

Cohen, G. L., Aronson, J., & Steele, C. M. (2000). When beliefs yield to evidence: Reducing biased evaluation by affirming the self. *Personality and Social Psychology Bulletin, 26,* 1151–1164.

Cooper, J., & Fazio, R. (1984). A new look at dissonance theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 17, pp. 229–268). San Diego, CA: Academic Press.

Czopp, A. M., Monteith, M. J., & Mark, A. Y. (2006). Standing up for change: Reducing bias through interpersonal confrontation. *Journal of Personality and Social Psychology, 90,* 784–803.

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

Elliot, A. J., & Devine, P. G. (1994). On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *Journal of Personality and Social Psychology, 67,* 382–394.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7,* 117–140.

Festinger, L. (1957). *A theory of cognitive dissonance.* New York: Harper & Row.

Frankena, W. K. (1973). *Ethics* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Fromm, E. (1941). *Escape from freedom.* New York: Rinehart.

Galinsky, A. D., Stone, J., & Cooper, J. (2000). The reinstatement of dissonance and psychological discomfort following failed affirmations. *European Journal of Social Psychology, 30,* 123–147.

Gosling, P., Denizeau, M., & Oberlé, D. (2006). Denial of responsibility: A new mode of dissonance reduction. *Journal of Personality and Social Psychology, 90,* 722–733.

Latané, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* Englewood Cliffs, NJ: Prentice Hall.

Maas, P. (1973). *Serpico: The cop who defied the system.* New York: Viking Press.

Major, B., Testa, M., & Bylsma, W. H. (1991). Responses to upward and downward social comparisons: The impact of esteem-relevance and perceived control. In J. Suls & T. A. Wills (Eds.), *Social comparison: Contemporary theory and research* (pp. 237–260). Hillsdale, NJ: Erlbaum.

Milgram, S. (1965). Liberating effects of group pressure. *Journal of Personality and Social Psychology, 1,* 127–134.

Milgram, S. (1974). *Obedience to authority*: *An experimental view.* New York: Harper & Row.

Monin, B. (2007). Holier than me? Threatening social comparison in the moral domain. *International Review of Social Psychology, 20*(1), 53–68.

Monin, B., Pizarro, D. A., & Beer, J. S. (2007). Deciding versus reacting: Conceptions of moral judgment and the reason-affect debate. *Review of General Psychology, 11,* 99–111.

My Lai massacre hero dies at 62. (2006, January 6). *BBC News*. Retrieved November 27, 2007, from http://news.bbc.co.uk/2/hi/americas/4589486.stm

Nietzsche, F. (1984). *Human, all too human: A book for free spirits.* Lincoln: University of Nebraska Press. (Original work published 1878)

Park, H., & Ybarra, O., & Stanik, C. (2007). *Self-judgment and reputation monitoring: Implications for finding stable and dynamic self-aspects.* Manuscript submitted for publication.

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

Rawls, J. (1999). *A theory of justice* (Rev. ed.). Cambridge, MA: Harvard University Press.

Rosin, H. (2004, May 17). When Joseph comes marching home: In a Western Maryland town, ambivalence about the son who blew the whistle at Abu Ghraib. *Washington Post*, p. C01.

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in *Personality and Social Psychology Bulletin*. *Personality and Social Psychology Bulletin, 28,* 1629–1646.

Sabini, J., & Silver, M. (1982). Moral reproach. In J. Sabini & M. Silver (Eds.), *Moralities of everyday life* (pp. 35–53). Oxford, England: Oxford University Press.

Salovey, P. (1991). Social comparison processes in envy and jealousy. In J. Suls & T. A. Wills (Eds.), *Social comparison: Contemporary theory and research* (pp. 261–285). Hillsdale, NJ: Erlbaum.

Sartre, J. P. (1956). *Being and nothingness: An essay on phenomenological ontology*. New York: Philosophical Library. (Original work published 1943)

Scher, S. J., & Cooper, J. (1989). Motivational basis of dissonance: The singular role of behavioral consequences. *Journal of Personality and Social Psychology, 56,* 899–906.

Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 38, pp. 183–242). New York: Academic Press.

Spencer, S. J., Fein, S., & Lomore, C. D. (2001). Maintaining one's self-image vis-à-vis others: The role of self-affirmation in the social evaluation of the self. *Motivation and Emotion, 25,* 41–65.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 261–302). New York: Academic Press.

Tangney, J. P., & Dearing, R. L. (2002). *Shame and guilt*. New York: Guilford Press.

Tesser, A. (1991). Emotion in social comparison and reflection processes. In J. Suls & T. A. Wills (Eds.), *Social comparison: Contemporary theory and research* (pp. 117–148). Hillsdale, NJ: Erlbaum.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, England: Cambridge University Press.

Walker, L. J., & Hennig, K. H. (2004). Differing conceptions of moral exemplarity: Just, brave, and caring. *Journal of Personality and Social Psychology, 86,* 629–647.

Ybarra, O. (2002). Naïve causal understanding of valenced behaviors and its implication for social information processing. *Psychological Bulletin, 128,* 421–441.

Zanna, M., & Cooper, J. (1974). Dissonance and the pill: An attribution approach to studying the arousal properties of dissonance. *Journal of Personality and Social Psychology, 29,* 703–709.

Appendix

Text of Tapes Used in Study 1

Obedient Condition

Reading week should be eliminated at [this university]. It just causes a number of problems. First, it breaks the momentum of the quarter, it also allows students less time to receive knowledge from the faculty, which is one of the reasons we're in school in the first place, right? We would learn more stuff with another week of classes. Other universities have more contact time with the professors, and it would make us more comparable with these other schools. These universities don't seem to be hurt by not having a reading week. Students tend to waste a lot of time during reading week, and they'd be better served by an extra week of classes. Also, because some of the schools at [this university] have a reading week and others don't, eliminating reading week would make the schedules equivalent across the schools at [this university].

Rebel Condition

So now I'm supposed to make a speech saying that reading week is a bad thing and that we should eliminate it, right? Well you know what? I don't think I'm going to do that. I know I was told to do it and I'm, like, a subject in the study but I'm still free to do whatever I want, right? And I'm not going to do something that I'm not 100% comfortable doing, like making the speech just because I'm told to. I'm sure that's my right as a participant in this experiment. I won't do it. There you have it—my official refusal. On tape and all. You can keep your money or whatever, I'd rather not get anything and not do something I have a problem with.