

# The Link Between Self-Dehumanization and Immoral Behavior



**Maryam Kouchaki, Kyle S. H. Dobson, Adam Waytz, and  
Nour S. Kteily**

Kellogg School of Management, Northwestern University

Psychological Science  
2018, Vol. 29(8) 1234–1246  
© The Author(s) 2018  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/0956797618760784  
www.psychologicalscience.org/PS



## Abstract

People perceive morality to be distinctively human, with immorality representing a lack of full humanness. In eight experiments, we examined the link between immorality and self-dehumanization, testing both (a) the causal role of immoral behavior on self-dehumanization and (b) the causal role of self-dehumanization on immoral behavior. Studies 1a to 1d showed that people feel less human after behaving immorally and that these effects were not driven by having a negative experience but were unique to experiences of immorality (Study 1d). Studies 2a to 2c showed that self-dehumanization can lead to immoral and antisocial behavior. Study 3 highlighted how self-dehumanization can sometimes produce downward spirals of immorality, demonstrating initial unethical behavior leading to self-dehumanization, which in turn promotes continued dishonesty. These results demonstrate a clear relationship between self-dehumanization and unethical behavior, and they extend previous theorizing on dehumanization.

## Keywords

morality, self-dehumanization, repeated dishonesty, open data, preregistered

Received 1/27/17; Revision accepted 1/28/18

Morality varies across cultures and individuals, but universally people consider morality to be a fundamental part of what makes us human (Brandt & Reyna, 2011; Haslam, Bastian, Laham, & Loughnan, 2011). Research on lay theories of human essence shows that people believe that moral sensibility constitutes a uniquely human property that distinguishes us from animals (Haslam, 2006). Indeed, denying someone uniquely human characteristics represents animalistic dehumanization that portrays people as “immoral or amoral (i.e., prone to violate the moral code or lacking it altogether)” (Haslam, 2006, p. 258). This view suggests a strong connection between immorality and dehumanization, with individuals seen as lacking morality deemed less human.

Here, we expand on the immorality-dehumanization link to examine the relationship between one’s own immoral behavior and self-dehumanization. Some work examining this question suggests that overtly immoral behavior in highly social contexts (e.g., ostracizing others) produces self-dehumanization and, as a result of reckoning with one’s immorality, subsequently increases moral behavior in an effort to restore one’s humanity

(Bastian et al., 2013). However, evidence also suggests that, in some subtler or more incremental contexts, initial immoral behavior can subsequently beget more immorality (Tenbrunsel & Messick, 2004). Consistent with this, the present research examined whether self-dehumanization in response to certain immoral acts increases subsequent immoral behavior. In particular, we examined whether immoral acts suggesting one’s lower capacities for agency (e.g., self-control) or experience (e.g., emotionality)—central dimensions of full humanity (Epley, Schroeder, & Waytz, 2013; H. M. Gray, Gray, & Wegner, 2007)—predict continued immoral behavior in succumbing to unethical temptation.

Self-dehumanization in response to initial unethical behavior might encourage subsequent immoral behavior for several reasons. For one, people might persist in their immoral behavior because, after their initial

## Corresponding Author:

Kyle S. H. Dobson, Northwestern University, Kellogg School of Management, 2211 Campus Dr., Evanston, IL 60208  
E-mail: kyle.dobson@kellogg.northwestern.edu

failures to avoid the temptation to behave dishonestly, they intuitively see themselves as lacking the agency and experience required for subsequent ethicality (Bem, 1972). Alternatively, perceiving that one lacks agency and experience after initial unethical behavior might help to subtly diminish one's culpability for their immoral actions, consistent with accounts of strategic self-deception (von Hippel & Trivers, 2011). Notably, Bastian et al.'s (2013) work examines contexts in which individuals have to explicitly grapple with the effects of their immoral behavior on other people (and its implications for their moral self-image). In contrast, our work presents an alternate paradigm in which immoral behavior is less explicitly social and in which behaving immorally may prompt people to subtly internalize an image of themselves as lacking the basic mental capacities required to resist the temptation to behave dishonestly, driving further immoral behavior.

The present research focused on a bidirectional pathway between immorality and dehumanization, testing both (a) the causal role of immoral behavior on self-dehumanization and (b) the causal role of self-dehumanization on immoral behavior. Importantly, if immoral behavior leads to self-dehumanization, and self-dehumanization can itself lead to immoral behavior, it may be the case that self-dehumanization can help partially explain downward spirals of dishonesty.

Only recently has work begun to examine self-dehumanization (Bastian & Haslam, 2010; Bastian et al., 2013; Yang, Jin, He, Fan, & Zhu, 2015) and suggest that self-dehumanization emerges among people who have been denied fundamental human needs, such as a connection with others (Baumeister & Leary, 1995) or control and power (Baumeister, 1999). Other work tied to the present research shows that, through a different mechanism, we sometimes self-dehumanize as a result of our own mistreatment of others. For example, research shows that people self-dehumanize (both in terms of traits considered essential to human nature and traits considered uniquely human) after showing aggression toward other players in a video game (Bastian, Jetten, & Radke, 2012) or after ostracizing other individuals (Bastian et al., 2013). Thus, observing oneself engaging in (presumably immoral) antisocial behavior may be enough to make an individual feel they are less than fully human. Immoral behavior, however, extends beyond aggression and ostracism. Indeed, the breadth and diversity of possible immoral behaviors (Haidt & Kesebir, 2010) necessitates investigating the link between immoral behavior and self-dehumanization beyond immoral behaviors rooted in directly harming identifiable others, such as social ostracism and aggression. We investigated this bidirectional relationship across a broader spectrum of behaviors, including lying,

cheating, and participants' self-generated recall of their own immoral acts.

More importantly still, how individuals respond to the state of self-dehumanization could differ on the basis of the unethical behavior that precedes it. Bastian et al. (2013) suggest (and demonstrate) that when people recognize that their "actions have caused harm to others that cannot be justified" (p. 157) as in the case of ostracism, they self-dehumanize and then behave more prosocially in an attempt to regain their humanity. Bastian et al. argue that in this sequence, taking personal responsibility for the immoral act is critical. However, considerable research also shows that initial immoral acts can lead to subsequent immoral acts (Bandura, 1999; Gino & Bazerman, 2009; Schrand & Zechman, 2012; Tenbrunsel & Messick, 2004)—in these contexts, repeated unethical behavior results from lowered awareness or ownership of one's morally questionable actions. Given the differences between our paradigm and that of Bastian et al. (2013), we expected and tested an alternate pattern in Study 3. Compared with their ostracism paradigm, our operationalization of unethical behavior lacks the social salience of a clear victim, does not ask participants to explicitly reflect on their own immorality or self-displeasure immediately after the initial unethical behavior, and represents a moral self-regulation failure in response to an immoral temptation. Because of these differences, we expected unethical behavior in this context to produce self-dehumanization that increases subsequent unethical behavior. This is particularly the case because our measure of self-dehumanization enables participants to excuse or explain their immoral behavior rather than take responsibility for it, therefore obviating the need to morally compensate.

## Overview of Studies

Studies 1a to 1c examined whether unethical (vs. ethical or neutral) behavior increases self-dehumanization. Study 1d compared unethical behavior (i.e., cheating on an exam) with a negative experience (i.e., failing an exam) to rule out effects of mere negativity. Studies 2a to 2c measured consequences of self-dehumanization on unethical behavior by directly manipulating self-dehumanization. Finally, to comprehensively test whether self-dehumanization can contribute to spirals of immorality, we conducted a preregistered study (Study 3) that provided participants with an initial opportunity to act dishonestly. In this study, we measured self-dehumanization as the result of their unethical act and then tested whether it predicted subsequent dishonest behavior.

We include all studies we conducted and, across all studies, we report all variables collected and all conditions included in the study designs. No participants

who completed our studies were excluded from the analyses unless noted for reasons identified prior to conducting the research. We examined the attrition rate and found no evidence of selective attrition across conditions (cf. Zhou & Fishbach, 2016). All data are available publicly at the Open Science Framework (<https://osf.io/ebhku/>).

## Study 1a

This study examined the effect of past unethical behavior on self-dehumanization. We compared unethical behavior with both ethical and neutral behavior.

## Method

**Participants and design.** One hundred seventy-seven individuals completed a short, online study on Amazon's Mechanical Turk website for \$0.50 each. We randomly assigned participants to three conditions (neutral, unethical act, ethical act). Given that Bastian et al. (2013) showed effect sizes ( $f$ s) from 0.40 to 0.73, we calculated our sample size on the basis of an estimated effect size of 0.40 (a large effect size). This required a sample size of approximately 150 participants for the study to achieve 99% power. We thus aimed for approximately 50 participants per condition (also consistent with the recommendations of Simmons, Nelson, & Simonsohn, 2013). Given the exclusion criteria developed prior to conducting the study, 21 participants who failed to follow instructions to write an essay or failed the attention check were excluded from analyses. The final sample thus consisted of 156 participants (68.6% male; age:  $M = 30.1$  years,  $SD = 10.1$ ; 49 in the unethical condition, 54 in the ethical condition, and 53 in the neutral condition).

**Procedure.** Participants were informed that they would complete several unrelated surveys. Participants were asked to recall an event and write about it for a few minutes, providing as many details as possible so that a person reading their entry would understand the situation and how they felt. We used the same instructions used in previous research (Barkan, Ayal, Gino, & Ariely, 2012). In the unethical condition, participants were asked to describe a situation where they did something unethical. In the ethical condition, they were instructed to write about something ethical from their past. In the neutral condition, they wrote about how they spend their evenings and described a typical instance.

We then measured self-dehumanization with 10 items (e.g., "How capable are you of doing things on purpose?" "How capable are you of experiencing emotion?"  $\alpha = .91$ ) from the Mind Attribution Scale, adapted from Kozak, Marsh, and Wegner (2006; 1 = *not at all*

*capable*, 7 = *extremely capable*). This measure of reduced mind attribution has been validated as a measure of dehumanization in previous work (K. Gray, Knobe, Sheskin, Bloom, & Barrett, 2011; Khamitov, Rotman, & Piazza, 2016; Waytz & Epley, 2012). As an attention check, we asked participants to indicate the topic of their writing task. In addition, we embedded an item on the self-dehumanization scale instructing them to select "3" for that item. They also completed a brief demographic questionnaire.

## Results

A one-way analysis of variance on self-dehumanization scores revealed a significant difference among the three conditions,  $F(2, 153) = 3.50$ ,  $p = .033$ ,  $\eta_p^2 = .04$ . Participants who recalled an unethical act self-reported marginally lower human capabilities ( $M = 5.82$ ,  $SD = 1.11$ ) compared with those in the neutral condition ( $M = 6.16$ ,  $SD = 0.73$ ,  $p = .064$ ) and significantly lower human capabilities compared with those who recalled an ethical act ( $M = 6.23$ ,  $SD = 0.65$ ,  $p = .022$ ). There was no significant difference between the ethical and neutral conditions ( $p = .61$ ; see Fig. 1).

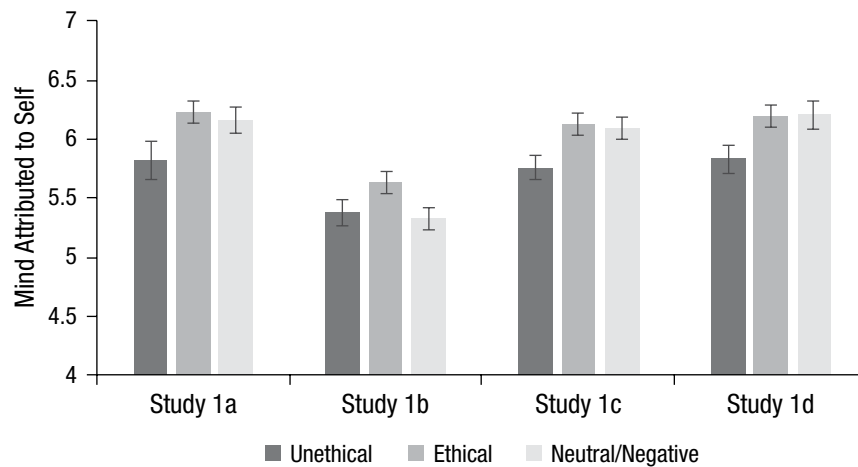
## Study 1b

Study 1a provided initial evidence for our theorizing. In Study 1b, we focused on instances of lying or telling the truth in the past 6 months to limit the variability in the type of immoral actions recalled. Additionally, we slightly modified our measure of dehumanization.

## Method

**Participants and design.** Two hundred fifty-five individuals completed a short, online study on Amazon's Mechanical Turk for \$0.50 each. We randomly assigned participants to three conditions (neutral, lying, honesty). As in Study 1a, we aimed for approximately 50 participants per condition. However, given that many participants in Study 1a did not write an essay or failed the attention check, we recruited more participants in this study to ensure a sufficient number. Given exclusion criteria developed prior to conducting the study (similar to Study 1a), 33 participants who failed to follow instructions to write an essay or failed the attention checks were excluded from analyses. The final sample thus consisted of 221 participants (62.9% male; age:  $M = 29.2$  years,  $SD = 9.9$ ; 69 in the unethical condition, 73 in the ethical condition, and 79 in the neutral condition).

**Procedure.** Participants were informed that they would complete several unrelated surveys. Participants were



**Fig. 1.** Mean rating of mind attributed to self across conditions (i.e., unethical, ethical, and neutral/negative) in Studies 1a to 1d. For Studies 1a to 1c, the third condition was “neutral,” while in Study 1d, the third condition was “negative.” Error bars represent  $\pm 1$  SEM.

asked to recall an event and write about it for a few minutes, providing as many details as possible so that a person reading their entry would understand the situation and how they felt. In the unethical condition, participants were asked to describe a time they behaved unethically by being dishonest and lying about something in the past 6 months. In the ethical condition, they were instructed to write about behaving ethically by being honest and telling the truth about something in the past 6 months. In the neutral condition, similar to Study 1a, they wrote about how they spend their evenings and described a typical instance.

We then measured self-dehumanization with a modified version of the 10 items from Study 1a, asking participants to compare themselves with an average person: (e.g., “Compared to the average person, how capable are you of doing things on purpose?” “Compared to the average person, how capable are you of experiencing emotion?”  $\alpha = .88$ ; 1 = *not at all capable*, 7 = *extremely capable*). As an attention check, we asked participants to indicate the topic of their writing task. In addition, we embedded an item on the self-dehumanization scale instructing them to select “3” for that item. They also completed a brief demographics questionnaire.

## Results

A one-way analysis of variance on self-dehumanization scores revealed a marginally significant difference among the three conditions,  $F(2, 218) = 2.67$ ,  $p = .071$ ,  $\eta_p^2 = .02$ . Participants who recalled lying self-reported marginally lower human capabilities ( $M = 5.38$ ,  $SD = 0.91$ ) compared with those who recalled telling the truth ( $M = 5.63$ ,  $SD = 0.78$ ,  $p = .077$ ). Unlike in Study 1a, we

found no significant difference between the unethical condition and the neutral condition ( $M = 5.33$ ,  $SD = 0.85$ ,  $p = .76$ ). However, the difference between the ethical and neutral conditions was significant ( $p = .026$ ; see Fig. 1). Despite the contrast between lying and honesty revealing a marginal effect, this study is broadly consistent with Study 1a in demonstrating that reminders of one’s own immorality tend to increase self-dehumanization compared with reminders of morality.

## Study 1c

One limitation of Studies 1a and 1b is that there was no experimental control over the details and nature of the specific situations people recalled. In Study 1c, we addressed this issue by randomly assigning people to read a story describing an ethical, unethical, or neutral situation using a first-person perspective-taking instruction.

## Method

**Participants and design.** Two hundred twenty-five students participated in the study for \$0.50 each. They were randomly assigned to three conditions (neutral, unethical act, ethical act). The sample size was determined by the number of participants who showed up during the scheduled sessions. Prior to conducting the study, we planned to stop data collection after the scheduled sessions were over, hoping to recruit at least 50 participants for each condition. This study was part of an hour-long series of studies. Five participants who failed attention checks were excluded from analyses. The final sample consisted of 220 participants (52.3% male; age:  $M = 26.6$  years,  $SD = 4.1$ ; 73 in the unethical condition, 75 in the ethical condition, and 72 in the neutral condition).

**Procedure.** Participants were asked to read a 200-word paragraph and then answer a few filler questions. In the neutral condition, they read a paragraph about wiki pages. In the ethical and unethical conditions, participants read a story from a first-person perspective and were instructed to put themselves in the situation (Kouchaki & Gino, 2016). In the story, participants imagined taking a chemistry course wherein they were struggling to get a passing grade. As a result, they made a cheat sheet as a backup for the final exam. During the exam, there was a question for which they tried for a few minutes to remember the correct answer. Participants in the unethical condition read that they could not remember the answer, so they used their hidden cheat sheet to write down the correct answer. Ultimately, they passed the course but felt very bad and guilty about cheating, believing they had done something morally wrong. In the ethical condition, participants read that after spending a couple of minutes thinking about the question, they remembered the answer so they did not use the hidden cheat sheet. They passed the course and felt very good and proud, believing they had done something morally right by not cheating.

In the subsequent survey, we used the same 10 items from Study 1a to measure self-dehumanization ( $\alpha = .91$ ). As an attention check, we asked participants to choose the topic of the essay they read about. In addition, we embedded an item on the self-dehumanization scale instructing them to select “3” for that item. As a manipulation check, we asked participants, at the end of the study, to think back to the essay they read and report how moral and ethical it made them feel on a 7-point scale (*extremely immoral/unethical* to *extremely moral/ethical*). They then completed a brief demographic questionnaire.

## Results

A one-way analysis of variance on each of the two manipulation questions revealed significant differences across conditions, both  $ps < .01$ . The unethical essay made participants feel less moral ( $M = 3.45$ ,  $SD = 0.96$ ) and ethical ( $M = 3.48$ ,  $SD = 1.08$ ) compared with both ethical (moral:  $M = 4.36$ ,  $SD = 0.80$ ; ethical:  $M = 4.03$ ,  $SD = 0.87$ ) and neutral (moral:  $M = 4.19$ ,  $SD = 0.87$ ; ethical:  $M = 3.92$ ,  $SD = 1.04$ ) essays.

A one-way analysis of variance with self-dehumanization scores as the dependent measure revealed a main effect of condition,  $F(2, 217) = 4.66$ ,  $p = .010$ ,  $\eta_p^2 = .04$ . Pairwise comparisons revealed that participants reported feeling a lower capacity for human attributes after reading the unethical story ( $M = 5.76$ ,  $SD = 0.88$ ) than after reading the ethical story ( $M = 6.13$ ,  $SD = 0.79$ ,  $p = .008$ ) or the neutral essay ( $M = 6.10$ ,  $SD = 0.78$ ,  $p = .014$ ). There was no significant difference between the ethical and neutral conditions ( $p = .85$ ; see Fig. 1).

## Study 1d

One concern with the previous studies is the possibility that recalling a negative or uncomfortable event rather than the unethicality of the act drove our results. Study 1d thus compared effects of unethical behavior with effects of a morally irrelevant negative experience.

## Method

**Participants and design.** One hundred fifty-four individuals completed a short, online study on Amazon's Mechanical Turk for \$0.50 each. They were randomly assigned to three conditions (ethical act, unethical act, negative event). As in Studies 1a to 1c, we aimed for approximately 50 participants per condition. One participant who failed attention checks was excluded from analyses. The final sample consisted of 153 participants (68.0% male; age:  $M = 30.5$  years,  $SD = 8.9$ ; 51 in the unethical condition, 52 in the ethical condition, and 50 in the negative condition).

**Procedure.** Participants were informed that they would complete several unrelated surveys. As in Study 1c, participants read a story and were asked to take a first-person perspective. We manipulated the unethical or ethical behavior as in Study 1c. In the negative-event condition, participants read a similar story in which they were struggling to get a passing grade in their chemistry course. As a result, they decided to hire a tutor to help them with the final exam (Kouchaki & Gino, 2016). Thus, there was no mention of the cheat sheet in this condition. During the exam, there was a question that they could not answer correctly. Ultimately, they did not pass the course and felt very bad because they failed the course.

We used the same 10 items from Studies 1a and 1c to measure self-dehumanization ( $\alpha = .90$ ). As an attention check, we asked participants to choose the topic of the essay they read about. In addition, we embedded an item on the self-dehumanization scale instructing them to select “3” for that item. As a manipulation check, we asked participants to think back to the essay they read and report the overall tone of the story (1 = *extremely negative*, 7 = *extremely positive*), how wrong and unethical the behavior was (*extremely wrong/unethical* to *extremely right/ethical*). They then completed a brief demographic questionnaire.

## Results

A one-way analysis of variance on each of three manipulation questions revealed significant differences across conditions, all  $ps < .001$ . The story about failing the test was perceived as more negative ( $M = 1.98$ ,  $SD = 0.82$ ) compared with both unethical ( $M = 2.43$ ,  $SD = 0.83$ )

and ethical ( $M = 5.00$ ,  $SD = 1.01$ ) stories. The unethical story was perceived as both more wrong and unethical (wrong:  $M = 2.94$ ,  $SD = 1.24$ ; unethical:  $M = 2.80$ ,  $SD = 1.41$ ) compared with both ethical (wrong:  $M = 4.42$ ,  $SD = 1.50$ ; unethical:  $M = 4.50$ ,  $SD = 1.51$ ) and negative (wrong:  $M = 4.40$ ,  $SD = 1.29$ ; unethical:  $M = 4.88$ ,  $SD = 1.55$ ) stories.

A one-way analysis of variance on self-dehumanization scores revealed significant differences across conditions,  $F(2, 150) = 3.61$ ,  $p = .029$ ,  $\eta_p^2 = .05$ . Pairwise comparisons revealed that participants reported feeling a lower capacity for human attributes after reading the unethical story ( $M = 5.83$ ,  $SD = 0.87$ ) than after reading the ethical story ( $M = 6.19$ ,  $SD = 0.69$ ,  $p = .023$ ) or the negative story ( $M = 6.21$ ,  $SD = 0.84$ ,  $p = .027$ ). There was no significant difference between the ethical and negative conditions ( $p = .87$ ; see Fig. 1).

### Single-Paper Meta-Analysis: Studies 1a to 1d

Given the slight variations in results and general consistency in design across Studies 1a to 1d, we conducted a single-paper meta-analysis (SPM) following procedures specified by McShane and Böckenholt (2017; see <http://www.singlepapermetaanalysis.com>). Across four experiments, the SPM showed that the unethical manipulation reduced mind attribution to oneself (i.e., increased self-dehumanization) compared with a control condition (neutral in Studies 1a to 1c, negative in Study 1d), effect =  $-0.34$ , 95% confidence interval (CI) =  $[-0.49, -0.19]$ , and compared with an ethical condition, effect =  $-0.24$ , 95% CI =  $[-0.39, -0.08]$ . No significant difference emerged between the neutral and ethical conditions, effect =  $0.10$ , 95% CI =  $[-0.03, 0.24]$ . Heterogeneity across methods of Studies 1a to 1d was very high, indicated by an  $I^2$  of 90% (95% CI =  $[84, 94]$ ), with method factors accounting for 90% of the total variation observed across the data beyond that attributable to the experimental manipulations. The CI width suggests that heterogeneity was estimated precisely.

This large heterogeneity prompted us to consider differences among the studies, and we determined that this high heterogeneity was virtually entirely attributable to Study 1b, which used a differently framed dependent measure, asking participants to rate themselves compared with the average person instead of asking them to rate themselves independently of others. Indeed, an SPM on the three studies not including Study 1b produced an  $I^2$  of 00% (95% CI =  $[0, 0]$ ), suggesting virtually no heterogeneity across Studies 1a, 1c, and 1d. Despite the methodological differences in Study 1b's design, the SPM across all four studies provides evidence that the two critical contrasts are significant, suggesting that unethicality leads to self-dehumanization.

## Study 2a

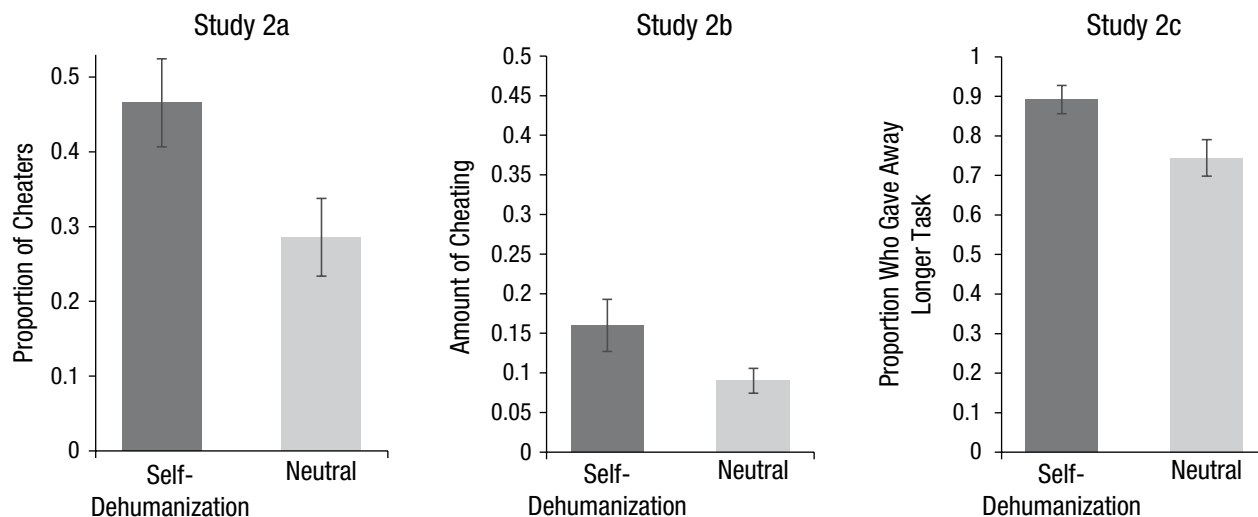
In Study 2a, we began examining the consequences of self-dehumanization by testing whether self-dehumanization increases dishonesty.

### Method

**Participants and design.** One hundred sixty individuals completed a short, online study on Amazon's Mechanical Turk for \$0.50 each and the opportunity to earn an additional \$1. We randomly assigned participants to two conditions (neutral, self-dehumanization). Given that we had no prior studies with which to estimate the effect size, we calculated our sample size on the basis of a medium effect size ( $f$ ) of 0.25. This required a sample size of approximately 80 participants per condition for a study powered at 85%. Ten participants who failed to follow instructions to write an essay or failed an attention check were excluded from analyses. The final sample consisted of 150 participants (54.7% male; age:  $M = 36.1$  years,  $SD = 12.7$ ; 73 in the self-dehumanized condition, and 77 in the neutral condition).

**Procedure.** Participants were informed that they would complete several unrelated surveys. To manipulate self-dehumanization, we used a writing task. We used the constructs of the scale in Studies 1a to 1d to instruct participants to think about a situation when they did not feel like themselves, with full capabilities as a human being. We informed them that other people engaging in this type of introspective task frequently write about instances in which they feel like they are not capable of exercising self-control, not capable of making plans and acting intentionally, not capable of remembering things well, not capable of experiencing feelings and emotion, or not capable of feeling pain or pleasure. In the neutral condition, we asked participants to write about their morning routines.

To measure immoral behavior as our dependent variable, we provided participants with a performance task. They were informed that they would be given 2 min to try to solve a series of four anagrams. Each set of letters could be unscrambled to spell an English word using all of the letters. They were informed that they would be paid on the basis of how many of them they unscrambled successfully without being asked to write out the unscrambled words. They received \$0.25 for each word they reported they had solved. Importantly, the fourth anagram was unsolvable, and we used the frequency with which participants reported having solved this last anagram as the measure of cheating (adapted from Kilduff, Galinsky, Gallo, & Reade, 2016). As an attention check, we embedded an item in the survey instructing them to select the fourth choice for



**Fig. 2.** Unethical behavior in the two conditions (i.e., self-dehumanization and neutral) in Studies 2a to 2c. Error bars represent  $\pm 1$  SEM.

that question. They also completed a brief demographic questionnaire.

## Results

As predicted, a higher percentage of participants in the self-dehumanized condition (46.6%, 34/73) than in the control condition (28.6%, 22/77) reported solving the unsolvable anagram,  $b = 0.78$ ,  $SE = 0.35$ ,  $OR = 2.18$ , Wald  $\chi^2(1, N = 150) = 5.11$ ,  $p = .024$  (see Fig. 2).

## Study 2b

To conceptually replicate Study 2a, we tested whether self-dehumanization increased a different form of dishonesty in Study 2b.

## Method

**Participants and design.** Two hundred individuals completed a short, online study on Amazon's Mechanical Turk for \$0.50 each and the opportunity to earn an additional \$0.50. They were randomly assigned to two conditions (neutral, self-dehumanization). As in Study 2a, we aimed for approximately 80 participants per condition. We recruited more participants to ensure a sufficient sample size given the possibility that participants might experience technical problems with the coin-flip task used in this study. Thirty-four participants were excluded because they failed to follow instructions to write an essay, failed the attention check, or had issues with the task (did not open the website, reported technical problems, or flipped the coin more than 10 times), thus making it impossible for us to verify whether they were

cheating or not. The final sample consisted of 166 participants (51.8% male; age:  $M = 34.3$  years,  $SD = 9.4$ ; 66 in the self-dehumanized condition, and 100 in the neutral condition). The unbalanced sample is due to experimenter error, with our web-based survey platform initially assigning 120 people to the neutral condition and 80 to the self-dehumanized condition. In our analyses, 20 participants in the neutral condition (17%) and 14 in the self-dehumanized condition (18%) were excluded on the basis of the exclusion criteria identified above.

**Procedure.** Participants were informed that they would complete several unrelated tasks. To manipulate self-dehumanization, we used the same writing tasks from Study 2a. We used a different cheating task to measure immoral behavior as our dependent variable. Participants were informed that they would complete a prediction task (adapted from Kouchaki & Gino, 2016) in which they would predict the outcome of a virtual coin flip for 10 rounds. In each round of the coin-flip task, they first indicated whether they expected heads or tails. Afterwards, they flipped a coin virtually and reported whether their prediction matched the actual outcome. They earned 5 cents if their prediction matched the outcome. Participants opened an ostensibly independent website that allowed them to virtually flip a coin. This website offered a true random coin flip, but unbeknownst to the participants, we recorded the outcome of each virtual coin flip and linked it to participants' identification numbers. Thus, we could identify whether participants cheated on any of the 10 rounds or not and compute a cheating ratio, dividing the number of overreports by the number of opportunities they had to overreport. A value of 1 indicates cheating to the fullest degree possible.

At the end, as a manipulation check, we asked participants to think back to the initial writing task they completed and indicate how it made them feel using the Mind Attribution Scale from previous studies ( $\alpha = .94$ ). As an attention check, we embedded an item in the dehumanization scale instructing them to select “2” for that item. They also completed a brief demographic questionnaire.

## Results

Participants who recalled an incident in which they lost their human capabilities indeed reported feeling a lower capacity for human attributes ( $M = 5.81$ ,  $SD = 0.94$ ) than those in the neutral condition ( $M = 6.23$ ,  $SD = 0.69$ ),  $F(1, 164) = 10.82$ ,  $p = .001$ ,  $\eta_p^2 = .06$ . Thus, our manipulation was successful.

A one-way analysis of variance on the average amount of cheating revealed a significant difference between the two conditions,  $F(1, 164) = 4.40$ ,  $p = .038$ ,  $\eta_p^2 = .02$ . Participants in the dehumanized condition cheated more ( $M = .16$ ,  $SD = .27$ ) compared with those in the neutral condition ( $M = .09$ ,  $SD = .16$ ; see Fig. 2).

## Study 2c

In Study 2c, we moved beyond cheating as an index of unethical behavior and tested our prediction that self-dehumanization would predict immoral behavior by manipulating self-dehumanization and then giving participants a choice of antisocial or prosocial behavior.

## Method

**Participants and design.** One hundred sixty-four individuals completed a short, online study on Amazon’s Mechanical Turk for \$0.50 each. We randomly assigned participants to two conditions (neutral, self-dehumanization). As in Studies 2a and 2b, we aimed for approximately 80 participants per condition. Four participants who failed to follow instructions to write an essay or failed an attention check were excluded from analyses. The final sample consisted of 160 participants (55.6% male; age:  $M = 35.2$ ,  $SD = 11.3$ ; 74 in the self-dehumanized condition and 86 in the neutral condition).

**Procedure.** Participants were informed that they would complete several unrelated tasks. To manipulate self-dehumanization, we used the same writing tasks from Studies 2a and 2b. For our dependent variable, participants were informed that for their next task, every person would complete a performance task. They further read that there were two tasks available and that they had been matched with another Mechanical Turk worker and

selected to assign tasks to themselves and this other person. They were given examples and detailed instructions about both tasks (a data entry task and a recognition task) and were assured that the choice was entirely up to them. In addition, we informed them that on the basis of responses from an earlier pilot study, the average time to complete the data entry task would be 3 min longer than for the recognition task. This task was based on previous research that operationalized immoral behavior on the basis of the decision of whether or not to assign another person a more burdensome task (Batson, Kobryniewicz, Dinnerstein, Kampf, & Wilson, 1997; Valdesolo & DeSteno, 2007). As an attention check, we asked participants to indicate which of the two tasks took longer to finish on average. They also completed a brief demographic questionnaire.

## Results

A higher percentage of participants in the self-dehumanized condition (89.2%, 66/74) than in the control condition (74.4%, 64/86) chose the longer task (data entry task) for the other worker,  $b = 1.04$ ,  $SE = 0.45$ ,  $OR = 2.84$ , Wald  $\chi^2(1, N = 160) = 5.40$ ,  $p = .020$  (see Fig. 2). This finding suggests that after self-dehumanizing, participants became more likely to engage in unethical (here, antisocial) behavior.

## Study 3

Study 3 tested a full model in which self-dehumanization might help to account for spirals of unethical behavior, with initial unethical behavior leading people to self-dehumanize and subsequently commit further unethical behavior. The original sample for this study was preregistered on the Open Science Framework (<https://osf.io/ebhku/>). After running the intended sample, we were encouraged by the editor to bolster our sample size. This additional data collection was also preregistered. The original sample and the additional sample were analyzed separately in our supplemental analyses, and we combined both samples in our main analyses and conducted sequential analyses. Because we added data to the original sample after having analyzed it, we report augmented  $p$  values in addition to standard  $p$  values, in line with recommendations for sequential analyses (Sagarin, Ambler, & Lee, 2014).

## Method

**Participants and design.** Four hundred twenty-nine students (33.0% male; age:  $M = 20.1$  years,  $SD = 2.0$ ) completed the study for \$12 each and the opportunity to earn an additional \$10. We randomly assigned participants to two conditions (neutral or possibility of cheating). In the



possibility-of-cheating condition (but not in the neutral condition), participants were confronted with a moral choice and could decide whether to be dishonest or not. We initially recruited 309 students and then recruited an additional 120.

**Procedure.** Participants were informed that they would complete several unrelated tasks. For the first task, participants were told that we wanted to assess their ability to predict the future. They were given one chance to predict what the result of a virtual coin flip would be and were told that they would earn \$2 if their prediction matched the outcome of the virtual coin flip. The coin flip was rigged so that the prediction of participants in the neutral condition was correct (i.e., their prediction matched the result); thus, they all earned \$2, which they were asked to take from the envelope placed on their desk. In the possibility-of-cheating condition, the result was rigged to always be inconsistent with the participant's prediction, but the page "mistakenly" reported to them that their prediction had been correct and they had thus earned \$2, which they were asked to take from the envelope placed on their desk. In all conditions, we gave participants a box asking, "Were there any problems? (SKIP this question if there were not any problems)." There was no reason for participants in the neutral condition to fill this box given that none experienced any problems.

After this future-prediction task, participants answered the same questions used in Studies 1a to 1d to measure self-dehumanization. Afterward, they completed a filler task of filling out "captchas" typically found on webpages for a couple of minutes. Finally, participants completed our dependent measure of unethical behavior (Task 2), which was a performance task similar to the Study 2a anagram task. For this task, participants had 2 min to solve eight anagrams, for which they would get \$1 for every anagram solved. The format was the same as Study 2a, except that three out of the eight anagrams were unsolvable. Thus, the measure of unethical behavior in Task 2 was the number of unsolvable anagrams reported solved. They could earn up to \$8 on the basis of their own self-report, which they were asked to take from the envelope placed on their desk.

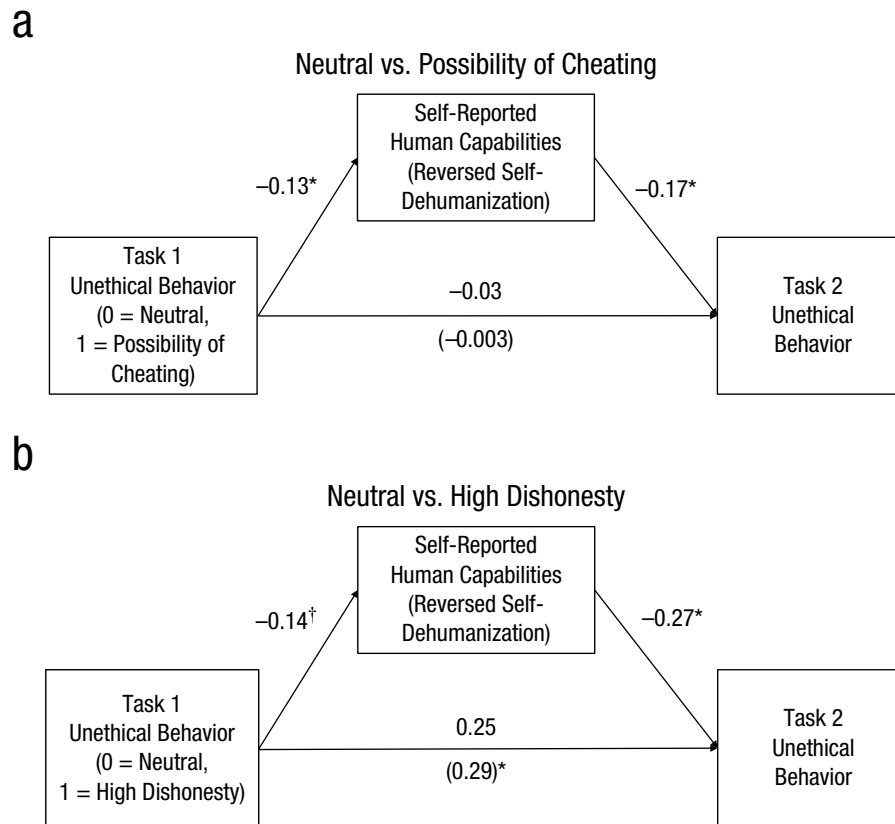
## Results

To test our hypothesis, we first examined the differences between the neutral and possibility-of-cheating conditions. A *t* test revealed a significant difference between these two conditions in self-dehumanization scores,  $t(427) = -2.11$ ,  $p = .036$ ,  $d = 0.22$ , augmented  $p = [.054, .065]$  (Sagarin et al., 2014). Participants in the

possibility-of-cheating condition self-reported significantly lower human capabilities ( $M = 5.83$ ,  $SD = 0.60$ ) compared with those in the neutral condition ( $M = 5.96$ ,  $SD = 0.57$ ). When examining the effect of the condition on Task 2 unethical behavior, we found no significant differences between conditions  $t(427) = 0.32$ ,  $p = .981$ , augmented  $p = [.961, .980]$ ; those in the possibility-of-cheating condition were not significantly more or less likely to cheat on the second measure of unethical behavior ( $M = 0.65$ ,  $SD = 1.04$ ) than those in the neutral condition ( $M = 0.65$ ,  $SD = 1.02$ ). To examine whether there was a significant indirect effect from initial cheating to self-dehumanization to subsequent cheating, we used the bootstrapping procedure recommended by Preacher, Rucker, and Hayes (2007). The 95% bias-corrected CI did not include zero, suggesting that the indirect effect was significant at the .05 level, indirect effect:  $b = 0.0222$ , 95% CI =  $[0.0011, 0.0705]$  (see Fig. 3a).

Next, and as detailed in our preregistration, we differentiated between participants in the possibility-of-cheating condition on the basis of whether or not they decided to honestly correct the apparent mistake and report the problem in the text box provided. We reasoned that more dishonest participants in the possibility-of-cheating condition would fail to report the error that had occurred, whereas less dishonest participants in the possibility-of-cheating condition would report that they were granted the bonus in error.

We coded participants' responses to the question asking them to report a problem. Of the 293 participants in the possibility-of-cheating condition, 159 participants reported a problem (i.e., low-dishonesty respondents), while 134 participants chose to remain silent and took the money they did not deserve (i.e., high-dishonesty respondents). We compared reported self-dehumanization between high-dishonesty participants and those in the neutral condition (i.e., excluding those who made the less unethical choice in the possibility-of-cheating condition). A *t* test revealed a marginally significant difference between conditions for self-dehumanization,  $t(268) = 1.96$ ,  $p = .051$ ,  $d = 0.24$ , augmented  $p = [.050, .076]$ . Those in the high-dishonesty condition self-reported marginally lower human capabilities ( $M = 5.82$ ,  $SD = 0.61$ ) compared with those in the neutral condition ( $M = 5.96$ ,  $SD = 0.57$ ). We then turned to unethical behavior in Task 2 to examine the effect of dishonesty in the coin-flip task on cheating behavior. We found significant differences between conditions  $t(262.21) = -2.14$ ,  $p = .033$ ,  $d = 0.26$ , augmented  $p = [.050, .064]$ . High-dishonesty participants were significantly more likely to cheat on the second measure of unethical behavior ( $M = 0.94$ ,  $SD = 1.17$ ) than participants in the neutral condition ( $M = 0.65$ ,  $SD = 1.02$ ). When examining whether self-dehumanization



**Fig. 3.** Mediation model showing the effect of Task 1 unethical behavior on Task 2 unethical behavior, as mediated by self-dehumanization (Study 3). Models are shown separately for the comparison between (a) the neutral and possibility-of-cheating conditions and (b) the neutral and high-dishonesty conditions. Unstandardized regression coefficients are shown. Along the bottom paths, the values in parentheses show the total effects, and the values outside parentheses show the direct effects after controlling for the mediator. Symbols indicate significant paths ( $^{\dagger}p = .051$ ,  $^*p < .05$ ).

mediated this path, we used the same bootstrapping procedure as in our previous analyses with the possibility-of-cheating conditions. As predicted, we found evidence of a significant indirect effect from initial cheating to subsequent cheating via self-dehumanization, indirect effect:  $b = 0.04$ , 95% CI = [0.0015, 0.1178] (see Fig. 3b). Overall, these findings support the idea that unethical behavior leads to self-dehumanization, which predicts more unethical behavior in the future.

## General Discussion

Eight experiments demonstrated morality's connection to humanness (i.e., mind attribution) and highlighted how self-dehumanization can affect subsequent moral behavior. Our effects varied in strength, with some results only marginally significant, yet a consistent pattern emerged across studies: Unethical behavior leads to subsequent self-dehumanization, and that self-dehumanization can in turn lead to downstream dishonesty.

These results contribute to the growing literature examining the consequences of engaging in unethical behavior and extend this work by examining the psychological mechanisms by which unethical behavior may be self-reinforcing. Our research both elucidates a consequence of unethical behavior—self-dehumanization—and demonstrates that this self-dehumanization can facilitate future unethical behavior. Interestingly, Bandura's seminal work on moral disengagement (Bandura, 1999; Bandura, Barbaranelli, Caprara, & Pastorelli, 1996) identifies dehumanization of other people as a means through which individuals justify immoral actions and that facilitates further immoral behavior. Here, instead, we identified dehumanization of oneself as a mechanism of moral disengagement that operates similarly.

Our research also expands the scope of research on self-dehumanization and immoral behavior. Prior work (Bastian et al., 2013) examining immorality in the social domain found evidence that self-dehumanization after

an initial immoral act (i.e., ostracizing another) led to subsequent compensatory moral behavior. In contrast, here we found that self-dehumanization following initial immorality reduces moral behavior. These differing patterns may result from the differing operationalizations of unethical behavior and dehumanization to which we alluded in the introduction. One notable distinction is that our work focused most broadly on self-regulation failures in the moral domain—that is “moral failings [that] are the result of weakness of the will” (Monin, Pizarro, & Beer, 2007, p. 105). Succumbing to the temptation to lie or cheat for money represents more of a self-regulatory failure than ostracizing another person, which does not involve such temptation.

Indeed, even though we suggest—like Bastian et al. (2013)—that immoral actions increase self-dehumanization, the moral self-regulatory failures we examined differed from their work in several important ways. Bastian et al. (2013) manipulated individuals' ostracism of others and then measured self-perceived immorality, global self-evaluation, and self-dehumanization. Evaluating one's immorality or self-displeasure could have induced a need to restore one's self-image in a way that our work did not. Bastian et al. (2013) also argue the self-dehumanization they examined “arises from the recognition that one's actions have caused harm to others that cannot be justified” (p. 157). They argue that people self-dehumanize in response to observing their own unjustified harm to others, while in our work, the immorality of the situation might not be as salient as the simple self-regulation failure of succumbing to a moral temptation. Another distinction is that Bastian et al.'s (2013) operationalizations of dehumanization focus on blatant dehumanization (e.g., referring to oneself as an animal), while we captured subtler self-dehumanization, in terms of recognition of one's reduced capacity for agency (e.g., planning, thought) and experience (e.g., emotion, feeling). In denying themselves human capacity, participants may excuse themselves from full responsibility for immoral behavior, and thus no moral compensation is necessary. Ultimately, we acknowledge that explanations for the differences between our work and prior work are largely speculative, and future research could differentiate between these findings more definitively. Nonetheless, our findings importantly demonstrate that self-dehumanization does not always lead to compensatory moral behavior and can sometimes contribute to downward spirals of immorality.

Future research could also investigate the precise mechanisms through which self-dehumanization causes unethical behavior. We observed that self-dehumanization significantly mediated the link between initial unethical behavior and repeated unethicality in Study 3. These

results clearly highlight the role that self-dehumanization plays in leading to downstream unethical behavior, providing a clear basis for future work further explicating the underlying pathways. We coded the specific content of the self-dehumanization essays in Studies 2a to 2c (see the Supplemental Material available online) to identify whether particular features of the self-dehumanization essays accounted for the effects. We found that self-dehumanization essays and control essays differed in several content areas. For example, our coding shows that when writing about an instance of feeling less than fully human, individuals' responses indicated lower capacities for agency including lower self-control (i.e., self-regulation is a known contributor to dishonesty; Gino, Schweitzer, Mead, & Ariely, 2011), more positive and negative emotion, and lower levels of capacities for experience (i.e., emotion, feeling). We also examined the associations between these constructs and cheating behavior. Although none of the constructs we coded for were consistently strongly associated with cheating (nor did any significantly mediate effects), our analyses suggest that the lack of agency reflected in these essays may contribute to subsequent unethicality (see the Supplemental Material). It is possible that our coding method is insensitive to capturing aspects of self-dehumanization that do not appear explicitly in participants' written essays (which tended to be short; word count:  $M = 42.0$ ,  $SD = 28.6$ ) and that emerge more subconsciously. Future work using alternate methods could systematically examine the specific role of self-perceived reduced agency and related constructs in explaining the link between self-dehumanization and subsequent unethicality.

Despite questions about underlying processes, the potential societal implications of our findings are clear. Our work suggests that people who face dehumanization and internalize it (e.g., prisoners, the homeless) may continue down a path of immoral behavior, making rehabilitative programs that emphasize restoration of one's feelings of humanity essential. Though many may believe that making a prisoner or criminal feel inhuman is necessary to punish them or motivate them to improve, this dehumanization might cause them to disconnect from morality as a fundamental facet of humanity. Eliminating self-dehumanization could drive people to believe they are good and human and motivate them to act the part. Importantly, this logic extends to other groups, such as politicians or corporate CEOs, that people may consider to be immoral and thus less human (e.g., Taibbi, 2009, in the aftermath of the financial crisis, referred to Goldman Sachs as a “vampire squid,” p. 52). Such depictions might lead these groups to internalize dehumanized perceptions, which may reinforce a path of immoral behavior rather than motivate them to correct course.

## Action Editor

Bill von Hippel served as action editor for this article.

## Author Contributions

All authors developed the study concept and contributed to the study design. Testing and data collection were performed by M. Kouchaki and K. S. H. Dobson, both of whom also analyzed and interpreted the data with A. Waytz and N. S. Kteily. M. Kouchaki, K. S. H. Dobson, and A. Waytz drafted the manuscript, and all authors provided critical revisions. All authors approved the final version of the manuscript for submission.

## Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

## Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797618760784>

## Open Practices



All data have been made publicly available via the Open Science Framework and can be accessed at <https://osf.io/ebhku/>. The design and analysis plan for Study 3 were pre-registered at <https://osf.io/ebhku/>. Materials for these studies have not been made publicly available. The complete Open Practices Disclosure for this article can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797618760784>. This article has received badges for Open Data and Preregistration. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.

## References

- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3, 193–209.
- Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology*, 71, 364–374.
- Barkan, R., Ayal, S., Gino, F., & Ariely, D. (2012). The pot calling the kettle black: Distancing response to ethical dissonance. *Journal of Experimental Psychology: General*, 141, 757–773.
- Bastian, B., & Haslam, N. (2010). Excluded from humanity: The dehumanizing effects of social ostracism. *Journal of Experimental Social Psychology*, 46, 107–113.
- Bastian, B., Jetten, J., Chen, H., Radke, H. R., Harding, J. F., & Fasoli, F. (2013). Losing our humanity: The self-dehumanizing consequences of social ostracism. *Personality and Social Psychology Bulletin*, 39, 156–169.
- Bastian, B., Jetten, J., & Radke, H. R. (2012). Cyber-dehumanization: Violent video game play diminishes our humanity. *Journal of Experimental Social Psychology*, 48, 486–491.
- Batson, C. D., Kobrynowicz, D., Dinnerstein, J. L., Kampf, H. C., & Wilson, A. D. (1997). In a very different voice: Unmasking moral hypocrisy. *Journal of Personality and Social Psychology*, 72, 1335–1348.
- Baumeister, R. F. (1999). *Evil: Inside human violence and cruelty*. New York, NY: Henry Holt.
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117, 497–529.
- Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 1–62). San Diego, CA: Academic Press.
- Brandt, M. J., & Reyna, C. (2011). The chain of being a hierarchy of morality. *Perspectives on Psychological Science*, 6, 428–446.
- Epley, N., Schroeder, J. R., & Waytz, A. (2013). Motivated mind perception: Treating pets as people and people as animals. In S. Gervais (Ed.), *Nebraska symposium on motivation* (Vol. 60, pp. 127–152). New York, NY: Springer.
- Gino, F., & Bazerman, M. H. (2009). When misconduct goes unnoticed: The acceptability of gradual erosion in others' unethical behavior. *Journal of Experimental Social Psychology*, 45, 708–719.
- Gino, F., Schweitzer, M. E., Mead, N. L., & Ariely, D. (2011). Unable to resist temptation: How self-control depletion promotes unethical behavior. *Organizational Behavior and Human Decision Processes*, 115, 191–203.
- Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, 315, 619.
- Gray, K., Knobe, J., Sheskin, M., Bloom, P., & Barrett, L. F. (2011). More than a body: Mind perception and the nature of objectification. *Journal of Personality and Social Psychology*, 101, 1207–1220.
- Haidt, J., & Ksebir, S. (2010). Morality. In S. Fiske, D. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., pp. 797–832). Hoboken, NJ: Wiley.
- Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review*, 10, 252–264.
- Haslam, N., Bastian, B., Laham, S., & Loughnan, S. (2011). Humanness, dehumanization, and moral psychology. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 203–218). Washington, DC: American Psychological Association.
- Khamitov, M., Rotman, J. D., & Piazza, J. (2016). Perceiving the agency of harmful agents: A test of dehumanization versus moral typecasting accounts. *Cognition*, 146, 33–47.
- Kilduff, G. J., Galinsky, A. D., Gallo, E., & Reade, J. J. (2016). Whatever it takes to win: Rivalry increases unethical behavior. *Academy of Management Journal*, 59, 1508–1534.
- Kouchaki, M., & Gino, F. (2016). Memories of unethical actions become obfuscated over time. *Proceedings of the National Academy of Sciences, USA*, 113, 6166–6171.
- Kozak, M. N., Marsh, A. A., & Wegner, D. M. (2006). What do I think you're doing? Action identification and mind

- attribution. *Journal of Personality and Social Psychology*, 90, 543–555.
- McShane, B. B., & Böckenholt, U. (2017). Single-paper meta-analysis: Benefits for study summary, theory testing, and replicability. *Journal of Consumer Research*, 43, 1048–1063.
- 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.
- Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, 42, 185–227.
- Sagarin, B. J., Ambler, J. K., & Lee, E. M. (2014). An ethical approach to peeking at data. *Perspectives on Psychological Science*, 9, 293–304.
- Schrand, C. M., & Zechman, S. L. (2012). Executive overconfidence and the slippery slope to financial misreporting. *Journal of Accounting & Economics*, 53, 311–329.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2013, January). *Life after P-hacking*. Paper presented at the Meeting of the Society for Personality and Social Psychology, New Orleans, Louisiana. Retrieved from [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2205186](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2205186)
- Taibbi, M. (2009, April). The great American bubble machine. *Rolling Stone*, 1082/1083, 52–101.
- Tenbrunsel, A. E., & Messick, D. M. (2004). Ethical fading: The role of self-deception in unethical behavior. *Social Justice Research*, 17, 223–236.
- Valdesolo, P., & DeSteno, D. (2007). Moral hypocrisy social groups and the flexibility of virtue. *Psychological Science*, 18, 689–690.
- von Hippel, W., & Trivers, R. (2011). Reflections on self-deception. *Behavioral and Brain Sciences*, 34, 41–56.
- Waytz, A., & Epley, N. (2012). Social connection enables dehumanization. *Journal of Experimental Social Psychology*, 48, 70–76.
- Yang, W., Jin, S., He, S., Fan, Q., & Zhu, Y. (2015). The impact of power on humanity: Self-dehumanization in powerlessness. *PLOS ONE*, 10(5), Article e0125721. doi:10.1371/journal.pone.0125721
- Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology*, 111, 493–504.