



Dehumanization increases instrumental violence, but not moral violence

Tage S. Rai^{a,1}, Piercarlo Valdesolo^b, and Jesse Graham^c

^aSloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139; ^bDepartment of Psychology, Claremont McKenna College, Claremont, CA 91711; and ^cDepartment of Psychology, University of Southern California, Los Angeles, CA 90007

Edited by Lee D. Ross, Stanford University, Stanford, CA, and approved June 22, 2017 (received for review April 5, 2017)

Across five experiments, we show that dehumanization—the act of perceiving victims as not completely human—increases instrumental, but not moral, violence. In attitude surveys, ascribing reduced capacities for cognitive, experiential, and emotional states to victims predicted support for practices where victims are harmed to achieve instrumental goals, including sweatshop labor, animal experimentation, and drone strikes that result in civilian casualties, but not practices where harm is perceived as morally righteous, including capital punishment, killing in war, and drone strikes that kill terrorists. In vignette experiments, using dehumanizing compared with humanizing language increased participants' willingness to harm strangers for money, but not participants' willingness to harm strangers for their immoral behavior. Participants also spontaneously dehumanized strangers when they imagined harming them for money, but not when they imagined harming them for their immoral behavior. Finally, participants humanized strangers who were low in humanity if they imagined harming them for immoral behavior, but not money, suggesting that morally motivated perpetrators may humanize victims to justify violence against them. Our findings indicate that dehumanization enables violence that perpetrators see as unethical, but instrumentally beneficial. In contrast, dehumanization does not contribute to moral violence because morally motivated perpetrators wish to harm complete human beings who are capable of deserving blame, experiencing suffering, and understanding its meaning.

moral | violence | dehumanization | instrumental | aggression

The failure to recognize other people as fellow human beings is considered to be a fundamental enabler of violence across cultures and throughout history (1–4). Known as dehumanization, this process is thought to have allowed colonists to exterminate indigenous peoples as if they were insects and whites to own blacks as if they were property (5–7). The dehumanization hypothesis developed amid theories arguing that violence is motivated primarily by instrumental gain or impulsive reactions, and that a greater appreciation of our shared humanity would lead to more peaceful relations (8–10). Although these theories differ in focus, collectively they assume that violence is restrained in part by a sense of moral obligation and sympathy toward fellow human beings, whom we feel are entitled to rights and protections that prohibit violence against them. In this framework, dehumanization causes perpetrators to perceive victims as nonhuman and, therefore, not entitled to moral obligation or sympathy, thus enabling perpetrators to act out their violent impulses free of inhibition and without remorse (11–15).

The basic premise on which the dehumanization hypothesis depends—that violence is restrained by moral inhibitions against harming fellow human beings—is in flux. Recent ethnographic and historical analyses, and classical works on the evolution of cooperation and the sociology of crime, indicate that descriptively, many perpetrators feel their violence is righteous and that their victims deserve what is coming to them. The husband who avenges the murder of his wife, the vigilante who cripples criminals, the soldier who kills the enemy, the public that votes for capital punishment, and even the suicide terrorist who detonates a bomb—all

may see violence as morally justified, obligatory, and even praiseworthy (16–23). Meanwhile, recent philosophical analyses have argued that the logic of dehumanization is inconsistent with perpetrators' actions. Specifically, the failure to recognize someone's humanity predicts indifference toward their suffering, not an active desire to bring that suffering about. Perpetrators should feel no need to humiliate, rape, and torture their victims, to force them to watch each other suffer, to conduct show trials, or to claim to their victims that the violence being done to them is morally laudatory (10, 24–26). At the same time, research on attribution and moral judgment has found that moral blame is directed toward people who plan and intend their actions, who have greater control and ability, and who understand why they are being harmed, all of which would seem to require recognition of human capacity (27–29). Even the American justice system imposes lighter sentences on individuals who lack these capacities, including minors, people with mental disabilities, and the criminally insane (30). Taken together, these findings indicate that our sense of morality does not only inhibit violence, it may also impel violence against victims recognized as fellow human beings. If so, the role of dehumanization in violence is unclear, as moral violence may not be driven by the weakening of moral inhibitions, but rather by the strengthening of moral motives.

Here, we theorize that morally motivated perpetrators wish to harm victims who deserve it, can experience it fully, and understand its meaning. To do so, their victims must be capable of thinking and having intentions, feeling pain and other sensations, and experiencing moral emotions—they must be human. Therefore, we hypothesize that perpetrators should feel no need to dehumanize

Significance

To eliminate violence, we must understand the motives that drive it. Most theories assume that violence is motivated by instrumental gain or impulsiveness, and is restrained by moral inhibitions. In these frameworks, dehumanization breaks down moral inhibitions by reducing perceptions of victims as fellow human beings worthy of concern. However, we argue that much violence is actually motivated by moral sentiments, and that morally motivated perpetrators wish to harm fellow human beings. Across five experiments, we show that dehumanizing victims increases instrumental, but not moral, violence. This distinction, between instrumental violence enabled by dehumanization, and moral violence directed toward human victims, has important implications for understanding how morality and dehumanization interact with violence, and for informing violence reduction efforts worldwide.

Author contributions: T.R., P.V., and J.G. designed research; T.R. performed research; P.V. and J.G. contributed new analytic tools; T.R. analyzed data; and T.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. Email: tage@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1705238114/-DCSupplemental.

victims when their violence is motivated by moral sentiments. If confirmed, the moral violence hypothesis would suggest the existence of two discrete forms of violence: one experienced as morally objectionable but nonetheless desirable for instrumental reasons (instrumental violence), and another motivated by moral sentiments that require violence despite any moral inhibitions against it (moral violence). According to this hypothesis, dehumanization enables instrumental violence by weakening moral inhibitions that would otherwise restrain it, thus making perpetrators apathetic to victims' suffering. In contrast, dehumanization cannot cause moral violence because it would strip victims of the qualities that necessitated the violence in the first place. Support for this hypothesis would demonstrate the importance of examining moral violence as a distinct form of violence that makes unique predictions and cannot be analyzed through the lens of previous theories.

Defining Dehumanization and Moral and Instrumental Violence

To explain violent practices, measures of dehumanization often combine the denial of human attributes with the addition of negatively valenced attributes. Thus, they operationalize dehumanization as the willingness to describe victims as animals, apes, and worms (31–33); as individuals who are superficial, cold, unsophisticated, and unrefined (34, 35); or as individuals who elicit contempt and disgust (36). The advantage of this approach is that it accounts for a wide range of violent phenomena. The disadvantage is that in attributing violent acts to dehumanization, these measures struggle to experimentally disentangle the removal of human attributes from the addition of negative attributes, or to establish dehumanization as a construct distinct from animosity or moral outrage (1). Rather than reflecting failed human perception, it is possible that when perpetrators describe victims in animalistic, disgusting, or other negative terms, their aim is to degrade and exert dominance over someone whom they know is human and wishes to be identified as such (10, 25, 26).

To overcome these limitations, we define dehumanization strictly as the failure to engage in social cognition of other human minds (37–39). From this perspective, what makes someone human is the existence of an “inner life” comprised of particular cognitive and emotional states and sensations (40, 41). Specifically, we define dehumanization in terms of capacities for agency (e.g., intending, planning, reasoning, remembering), experience (e.g., pain, hunger, fear, pleasure), and counterbalanced positive and negative moral emotions (e.g., love, compassion, anger, hatred) (42, 43). In line with previous research, we do not directly measure ascriptions of “humanness,” as such abstract, global measures have been found to be more susceptible to contextual biases and demand characteristics than measures of specific attributes (44). This approach may not capture all violent acts to which the term “dehumanization” has colloquially been applied. Indeed, we theorize that many violent practices that have been attributed to dehumanization are actually driven by different psychological mechanisms. In addition, it is possible that the human “essence” compromised by dehumanization cannot be assessed through measures of specific human attributes (45). However, to the extent that measures of specific human attributes do capture dehumanization, our approach isolates the denial of human attributes from the addition of negative attributes and provides a theoretically grounded scientific definition from which to identify the distinct psychological mechanisms that give rise to violence.

Regarding our distinction between moral and instrumental violence, there is considerable disagreement as to what constitutes a “moral” judgment and whether it has any unifying features (46). In addition, some violent acts may be characterized by both moral and instrumental features. In our experiments, instrumental violence is intended to refer to cases that are primarily characterized by perpetrators who do not necessarily desire to harm victims, but who knowingly harm them in order to achieve some other objective.

Meanwhile, moral violence is intended to refer to cases that are primarily characterized by perpetrators who harm victims because they see them as deserving of it.

Overview of Experiments

To test our hypothesis that dehumanization increases instrumental, but not moral, violence, we conducted five experiments. Experiments 1 and 2 use attitude surveys to investigate whether natural variation in dehumanizing attitudes toward victims predicts support for instrumental, but not moral, violence in political and intergroup contexts. Experiment 3 uses a vignette-based experimental design to provide causal evidence for our hypothesis by investigating whether describing strangers in dehumanized terms increases willingness to harm them for instrumental, but not moral, reasons. Experiment 4 examines the reverse causal pathway by investigating whether people spontaneously dehumanize strangers when they imagine harming them for instrumental, but not moral, reasons. Finally, Experiment 5 extends our hypothesis by investigating whether people spontaneously humanize strangers who are severely lacking in human attributes when they imagine harming them for moral, but not instrumental, reasons.

Experiment 1

In Experiment 1 ($n = 187$), we drew on previous research on active and passive aggression (47) to present participants with either two instrumental violence practices or two moral violence practices. Participants in the instrumental condition were asked to state their level of approval for purchasing goods made in sweatshops and the use of animals in experiments (Fig. 1A). In these cases, we theorized that people instrumentally benefit from, but do not actively desire to harm, sweatshop laborers and laboratory animals. Dehumanization should predict greater support for these instrumental violence practices. Participants in the moral conditions were asked to state their level of approval for capital punishment of murderers and killing enemy soldiers in war. Although these cases have some instrumental features, we theorized that people actively desire to harm murderers and enemy soldiers because they feel these groups morally deserve it. Dehumanization should be unrelated to these moral violence practices. After participants expressed their level of approval, we measured dehumanizing attitudes by asking participants to assess victims' abilities to have intentions and make plans (agency), to experience pain and suffering (experience), to feel love and compassion, and anger and hate.

Perceiving less humanity in sweatshop laborers and laboratory animals predicted greater approval for the use of sweatshop labor and experiments on animals ($r = -0.39$, $P < 0.001$). In contrast, perceiving less humanity in murderers and enemy soldiers did not predict approval for capital punishment of murderers and killing in war ($r = -0.01$, $P = 0.887$). The two relationships were significantly different from each other ($z = 2.65$, $P = 0.008$). Support for instrumental violence was significantly negatively correlated with all four mental state measures, including agency ($r = -0.41$, $P < 0.001$), experience ($r = -0.35$, $P < 0.001$), love and compassion ($r = -0.29$, $P = 0.004$), and anger and hatred ($r = -0.27$, $P = 0.009$). In contrast, support for moral violence was not related to any of the individual mental state measures. Participants expressed greater approval for capital punishment and killing in war ($M = 4.25$, $SD = 1.70$) than for sweatshop labor and animal experimentation ($M = 2.65$, $SD = 1.19$; $t = 7.45$, $P < 0.001$), and they ascribed marginally greater humanity toward murderers and enemy soldiers ($M = 4.62$, $SD = 0.49$) than toward sweatshop laborers and animals ($M = 4.48$, $SD = 0.58$; $t = 1.85$, $P = 0.066$).

Experiment 2

Whereas the results of Experiment 1 support the moral violence hypothesis, the violent practices examined in the experiment vary on several dimensions, including the nature of the violence being

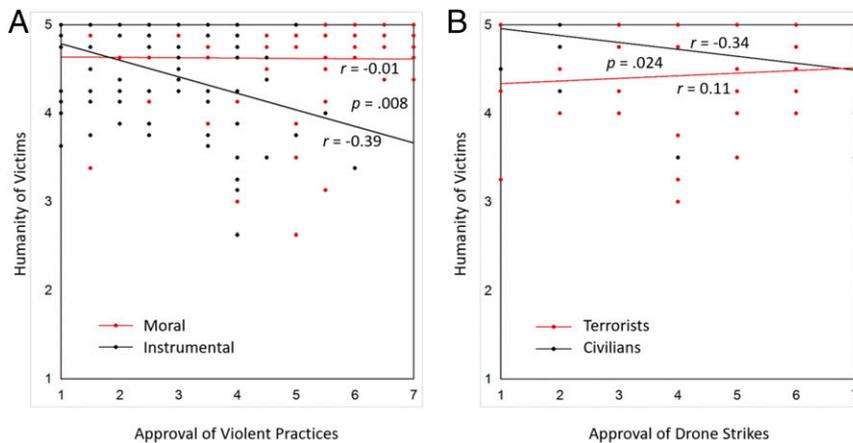


Fig. 1. Correlation plots displaying approval of violence and humanity of victims. (A) Approval of sweatshop labor and animal experimentation (instrumental violence) and approval of killing enemy soldiers in war and executing murderers (moral violence). (B) Approval of drone strikes framed around deaths to civilians and around deaths to terrorists. Each data point may represent several participants. Color of data points represent the type of violence the majority of those participants were responding.

committed. In addition, it was never confirmed whether participants viewed the violent acts as differing on instrumental and moral dimensions as theorized. In contrast, in Experiment 2 ($n = 101$), we assessed support for a single violent practice—military drone strikes—while framing the practice either around recipients of instrumental violence or recipients of moral violence (Fig. 1B). In the instrumental condition, we framed drone strikes around collateral damage to Iraqi civilians, and then measured support for drone strikes and humanity ascribed to Iraqi civilians. Although drone strikes may be perceived as utilitarian action directed toward a moral good of defeating terrorism, the Iraqi civilians are innocent and undeserving of harm, and so we theorized that in relation to the civilians specifically, the action is instrumental rather than moral. If participants do not actively wish to kill Iraqi civilians, but require them to die as a means to killing terrorists, then dehumanization should predict support for drone strikes. In the moral condition, we framed drone strikes around deaths to Islamic State of Iraq and the Levant (ISIS) terrorists, and then measured support for drone strikes and humanity ascribed to terrorists. If participants actively desire to harm terrorists because they feel they morally deserve it, then dehumanization should be unrelated to support for drone strikes.

We first conducted a pilot study to confirm that participants view drone strikes framed around deaths to ISIS terrorists as more strongly reflecting moral violence and drone strikes framed around collateral damage to Iraqi civilians as more strongly reflecting instrumental violence. We found that participants ($n = 96$) rated drone strikes framed around deaths to ISIS terrorists as significantly more moral ($M = 3.95$, $SD = 1.16$) than instrumental ($M = 2.65$, $SD = 1.35$; $t = 6.22$, $P < 0.001$), while they rated drone strikes framed around collateral damage to Iraqi civilians as significantly more instrumental ($M = 4.09$, $SD = 0.87$) than moral ($M = 2.14$, $SD = 1.18$; $t = 11.98$, $P < 0.001$).

Regarding the main results from Experiment 2, perceiving less humanity in Iraqi civilians predicted more approval for drone strikes ($r = -0.34$, $P = 0.016$). In contrast, perceiving less humanity in terrorists did not predict approval for drone strikes ($r = 0.11$, $P = 0.452$). The two relationships were significantly different from each other ($z = 2.25$, $P = 0.024$). Consistent with prior studies that manipulate dehumanization by varying moral behavior (44, 45), Iraqi civilians were seen as more human overall ($M = 4.83$, $SD = 0.42$) than ISIS terrorists ($M = 4.45$, $SD = 0.51$; $t = 4.05$, $P < 0.001$). This effect was driven by greater perceptions of capacities for love and compassion in Iraqi civilians ($M = 4.78$, $SD = 0.74$) than in ISIS terrorists ($M = 3.65$, $SD = 1.23$; $t = 5.60$, $P < 0.001$). However, support for drone strikes framed around deaths to ISIS terrorists was not related to dehumanization along any dimension, including love and compassion, and was marginally positively correlated with capacities for anger and hate ($r = 0.25$, $P = 0.077$). In contrast,

support for drone strikes framed around deaths to Iraqi civilians was significantly negatively correlated with perceptions of capacities for experience ($r = -0.44$, $P = 0.001$) and love and compassion ($r = -0.41$, $P = 0.003$). As would be expected, participants expressed greater approval for drone strikes when framed around deaths to terrorists ($M = 4.69$, $SD = 1.82$) than when framed around deaths to civilians ($M = 2.72$, $SD = 1.86$; $t = 7.45$, $P < 0.001$).

Experiment 3

Experiments 1 and 2 used natural variation in dehumanization to demonstrate that dehumanization is related to instrumental, but not moral, violence. Although these experiments provide external validity for the moral violence hypothesis, they do not provide causal evidence for it. Also, the higher approval of moral violence compared with instrumental violence may raise concerns. In Experiments 3–5, we use hypothetical vignette-based experiments to manipulate levels of dehumanization and motives to aggress, and more closely equate support for instrumental and moral violence.

In Experiment 3 ($n = 363$), we experimentally manipulated dehumanization and motives to aggress by describing a stranger in humanized or dehumanized terms and then assessing participants' willingness to harm the stranger for instrumental or moral reasons in a hypothetical scenario (Fig. 2). Participants were presented with a vignette in which they were told to imagine they had the opportunity to anonymously break a stranger's thumb. Participants' motives were manipulated by either telling participants that they would be paid \$2 million to do it (instrumental motive), or by telling participants that the stranger recruited poor young women into prostitution (moral motive). To manipulate dehumanization, we drew on previous research that identified personality traits that are seen as highly humanizing, but which are counterbalanced for positive and negative valence. Specifically, in the humanized condition, the stranger was described as a 29-year-old man with brown hair and brown eyes named John, who is "ambitious and imaginative, but also high-strung and insecure", whereas in the dehumanized condition, the stranger was simply described as a "man" (48, 49). Participants then rated their willingness to break the stranger's thumb and assessed the stranger's mental states.

Participants were significantly more likely to report that they would break the stranger's thumb for money when he was described in dehumanized ($M = 4.64$, $SD = 2.14$) rather than humanized terms ($M = 3.90$, $SD = 2.36$; $t = 2.26$, $P = 0.025$). In contrast, participants were no more likely to report that they would break the stranger's thumb for immoral behavior when he was described in dehumanized ($M = 4.04$, $SD = 2.37$) rather than humanized terms ($M = 4.13$, $SD = 2.24$; $t = 0.245$, $P = 0.807$). A two-way ANOVA analysis revealed a marginally significant interaction [$F(1, 360) = 2.97$, $P = 0.086$, $\eta^2 = 0.008$].

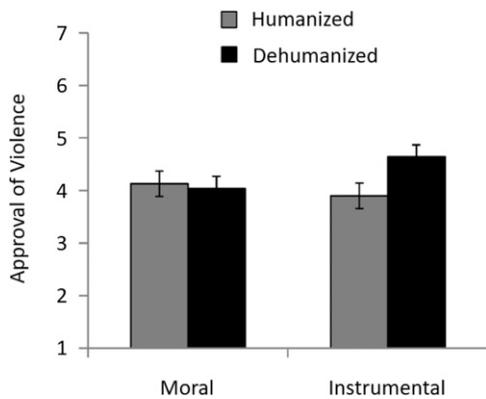


Fig. 2. Mean rating of willingness to harm a stranger as a function of instrumental or moral motive and whether the stranger was described in humanizing or dehumanizing language. Error bars represent one SEM.

Experiment 4

Experiment 3 provided causal evidence that when strangers are dehumanized, participants are more likely to support instrumental, but not moral, violence against them. In Experiment 4 ($n = 352$), we examined the reverse causal pathway by investigating whether consideration of violence to strangers for instrumental, but not moral, reasons would lead participants to spontaneously dehumanize them (Fig. 3A). Participants in two experimental conditions were presented with a vignette in which they were asked to imagine harming a stranger either for money or for immoral behavior. Participants in two control conditions were provided with matched vignettes, but were not asked to imagine harm. After reading the vignettes, all participants assessed the stranger's humanity.

Participants who imagined harming a stranger for money ($M = 3.91$, $SD = 0.87$) perceived the stranger as significantly less human than participants who did not imagine harming the stranger for money ($M = 4.46$, $SD = 0.63$; $t = 5.05$, $P < 0.001$). In contrast, participants who imagined harming an immoral stranger ($M = 3.98$, $SD = 0.80$) perceived the stranger as marginally more human than participants who did not imagine harming the immoral stranger ($M = 3.78$, $SD = 0.67$; $t = 1.75$, $P = 0.082$). A two-way ANOVA analysis revealed a significant interaction [$F(1, 349) = 22.59$, $P < 0.001$, $\eta^2 = 0.061$].

Experiment 5

Experiments 1–4 demonstrated that dehumanization predicts, causes, and is caused by instrumental, but not moral, violence. In the final experiment, we extend our thesis to hypothesize that when victims have strongly reduced human attributes, morally motivated perpetrators may be motivated to humanize victims to make their moral violence meaningful (23). To do so, we drew on previous research that found that people are ascribed reduced human attributes when they are unconscious (50). Thus, in Experiment 5 ($n = 362$), we presented participants with the same experimental design from Experiment 4, except we told them that the stranger whom they were considering harming was asleep (Fig. 3B). We predicted that when imagining harming the sleeping stranger for money, participants would feel no motivation to dehumanize the stranger further if the stranger was already lacking in humanity. In contrast, when imagining harming the sleeping stranger for immoral behavior, they would humanize the stranger to imbue the stranger with the humanity necessary to make moral violence meaningful.

No difference in perceived humanity was found between participants who imagined harming the stranger for money ($M = 2.99$, $SD = 1.04$) and participants who did not imagine harming the

stranger for money ($M = 2.85$, $SD = 0.99$; $t = 0.976$, $P = 0.331$). In contrast, participants who imagined harming the immoral stranger ($M = 3.60$, $SD = 0.95$) perceived the stranger as significantly more human than participants who did not imagine harming the immoral stranger ($M = 2.98$, $SD = 0.89$; $t = 4.50$, $P < 0.001$). A two-way ANOVA analysis revealed a significant interaction [$F(1, 359) = 5.48$, $P = 0.020$, $\eta^2 = 0.015$]. These results indicate that when perceived humanity is especially low, consideration of moral, but not instrumental, violence causes humanization of victims. Given that Experiments 1–4 found that instrumental violence was linked to dehumanization, whereas moral violence was not linked to any significant effects, one possibility is that there is a threshold effect to dehumanization, such that once victims are seen as sufficiently human, further humanization is unnecessary to sustain or generate moral violence, whereas once victims are sufficiently dehumanized, further dehumanization is unnecessary to sustain or generate instrumental violence.

Discussion

Harming another human being can be extraordinarily difficult and traumatic, but sometimes necessary or advantageous. One path to successfully committing violence is to disengage our moral inhibitions by dehumanizing victims, so that it no longer feels distressing to inflict pain on others (4, 45). However, a second path to successfully committing violence is to strengthen our moral motives to harm so that we can overcome, rather than disengage, any aversion to violence. Across five experiments, we provide external validity and causal evidence to indicate that dehumanization increases violence committed for instrumental, but not moral, reasons, and that consideration of instrumental, but not moral, violence causes dehumanization of victims. Morally motivated perpetrators may even humanize victims when it is necessary to generate moral meaning for the violence they do. The failure to recognize victims as fellow human beings does not make people desire to aggress, it simply makes them apathetic to victims' suffering when committing violence in pursuit of instrumental ends. When in pursuit of moral ends however, dehumanizing victims does not lead to violence, nor does violence lead to dehumanization, because dehumanization removes the very qualities that make moral violence meaningful.

These findings seem to run counter to the literature that has linked dehumanization to willingness to punish, torture, rape, and kill out-groups, women, and minorities (51–54). In this tradition, even our enemies and other “thin” relations must be perceived as human to prevent excessive violence toward them (55). Methodologically, the key differences between our findings and previous experimental results is that we operationalized dehumanization in terms of perceiving victims' minds rather than attributing animalistic or negative characteristics to them, and we distinguished moral from instrumental violence. The distinction between moral and instrumental violence has surface similarities with the dual model of dehumanization that distinguishes between “mechanistic” dehumanization, which is thought to cause instrumental violence in interpersonal contexts, and “animalistic” dehumanization, which is thought to cause “antagonistic” violence in intergroup contexts (2). However, that model would predict that instrumental and moral violence are driven by dehumanization along different mental state dimensions in different contexts. Instead, we found that dehumanization never predicted moral violence, it never caused moral violence, and it was never caused by moral violence in interpersonal or intergroup contexts.

One potential limitation of our method is that in practice, when perpetrators compare victims to animals, they may not only fail to perceive their victims' minds, they may imbue them with savage and contaminating attributes that are insulting and threatening to the perpetrators' sense of social hierarchy and in-group essence (1, 2). Perpetrators may also be morally conflicted or motivated by moral and instrumental reasons simultaneously. They may also engage in selective dehumanization, denying some human attributes while

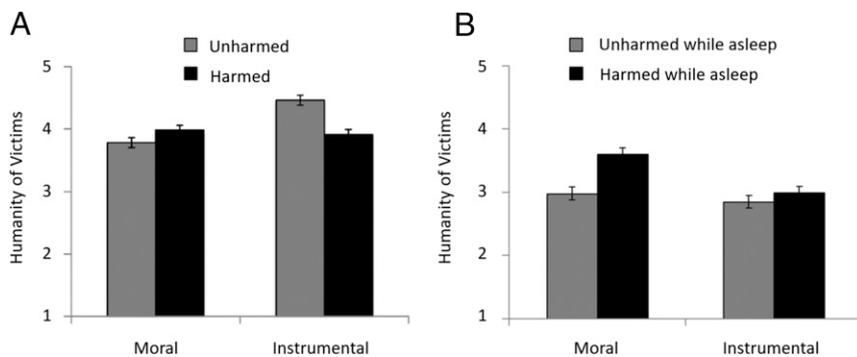


Fig. 3. Mean rating of the humanity of strangers as a function of motive and whether participants imagined harming the stranger. (A) When the stranger is awake. (B) When the stranger is asleep. Error bars represent one SEM.

upholding others. Each of these considerations raises issues regarding how researchers should conceptualize dehumanization moving forward. Nevertheless, our research reveals the existence of two distinct psychological processes tied to two discrete kinds of violence—in the first case, removing human attributes so that perpetrators and observers become apathetic to victims' welfare and can proceed with instrumental violence unimpeded by guilt or empathy, and in the second case, generating feelings of hostility and moral impetus to harm directly.

Our findings are as important as moral violence is prevalent. Many theories of moral psychology have argued that prohibitions against intentional harm form the core of our sense of morality, and incidents of violence can only be explained as undesirable but instrumentally necessary, or as failures to live up to our moral obligations (56–58). These moral psychological theories are mutually supported by most theories of violence, which assume that violence is primarily pursued for instrumental or impulsive reasons and restrained by our sense of moral obligation. Even when these theories consider “retaliatory” or “justice” motives, they are seen as either impulsive actions that perpetrators would not support upon reflection or as additional forms of utility that should not exhibit unique patterns (8, 59). Dehumanization has bound these various approaches together by arguing that violence is enabled when we fail to perceive our victims as fellow human beings worthy of moral concern.

However, ethnographic and historical evidence suggest otherwise. Police who abuse suspects, gangs that violently protect their turf, terrorists who commit suicide bombings, brothers who kill their sisters to cleanse family dishonor—all see their violence as morally justified, obligatory, and even virtuous (17). The pervasiveness of moral violence indicates that our findings are important, and in turn, our findings provide evidence that moral violence is fundamental. Its patterns cannot be understood or accurately predicted by theories that focus on prohibitions against harm, instrumental or impulsive motives for violence, or the failure to recognize the humanity of victims. Instead, our findings suggest that paradigms of moral psychology, violence, and dehumanization must shift toward a conceptualization of violence wherein its aim is often to intentionally aggress against complete human beings capable of deserving, suffering, and understanding the harm done to them.

In our new framework, combating dehumanization is still critical to reducing violence, but in a different manner than previously thought. Our findings suggest that when the Nazis compared Jews to rats in World War II, its primary effect may not have been to fuel the hatred of Nazi officials, but rather to reduce the sympathy felt by the German public toward the Jews being persecuted. Thus, dehumanization makes possible the everyday violence and large-scale atrocities that observers enable through their indifference. It allows us to refuse to help and to sacrifice others for the greater good because we are apathetic to their suffering (60, 61). It undermines reconciliation following conflict because there is no point to rebuilding relationships with those we think of as animals (55,

62). It also still enables the perpetuation of atrocities committed for instrumental gain through colonization, slavery, and other practices.

However, it does not cause morally motivated violence directly. At a time when Americans are enacting a ban on immigration from several Muslim majority countries, our findings suggest that there are two distinct psychological mechanisms at work. What enables the impulse is a mechanism of neglect caused by dehumanization of Muslim refugees who are mentally conceptualized as less than human. However, what drives the impulse is a mechanism of moral outrage directed toward terrorists who may be conceptualized as completely human. As Manne (10) notes in her description of Elliot Rodgers' mass shooting in Isla Vista, CA, in 2014, he claimed that he was punishing sorority women for not noticing him and “throwing themselves” at other men as part of his “revenge against humanity.” His awareness of women's capacities to think, feel, love, and choose did not defuse his outrage—it created it. The current research demonstrates that if we ever wish to fully understand violence, we cannot only study the breakdown of mechanisms that promote peace, we must also examine the moral motives that impel us to aggress in the first place.

Materials and Methods

Participants in all experiments were recruited via the Internet and compensated with \$0.25 following completion of a questionnaire administered through the Mechanical Turk site run by Amazon.com. It has been found that data collected from Amazon's Mechanical Turk site is as reliable as data gathered through traditional methods (63). All participants were drawn from the United States. The IP addresses of participants' computers were recorded to ensure that they did not participate in the study multiple times. Participants in all experiments were assigned to conditions randomly. Participants in all experiments reported demographic information including their political orientation, age, sex, ethnicity, and education level. Although women expressed less support for violence than men overall, no consistent meaningful interactions were predicted or found between the demographic variables and analyses of interest, and so those analyses are not reported in the paper. Participants in all experiments were also asked to guess our hypotheses. None succeeded. Participants in all experiments began by reading a short prompt about violence (*SI Appendix*). The Institutional Review Board at Northwestern University's Office for Research approved all research conducted in these experiments. Informed consent was obtained from all participants before participation in the experiments.

Experiments 1 and 2 used between-subjects designs. Participants first rated their approval of violent practices and then assessed humanity. Humanity items were rated on five-point likert scales ranging from 1 (not capable at all) to 5 (completely capable) and were counterbalanced for order (*SI Appendix*). In Experiment 1, participants in the “moral violence” condition reported the extent to which they approved of “executing murderers” and “killing enemy soldiers in a time of war” and assessed the humanity of “murderers” and “enemy soldiers,” whereas participants in the “instrumental violence” condition reported the extent to which they approved of “purchasing products that are made in so-called ‘sweatshops’ overseas” and “conducting experiments on animals,” and assessed the humanity of “people who work in sweatshops” and “animals.” In Experiment 2, participants in the “moral violence” condition reported the extent to which they approved of “drone strikes that result in the deaths of ISIS terrorists” and assessed the humanity of “ISIS terrorists,” whereas participants in the “instrumental violence” condition reported the extent to

which they approved of “drone strikes that result in collateral damage to innocent Iraqi civilians” and assessed the humanity of “Iraqi civilians.”

In the pilot study for Experiment 2, participants were presented with a distinction between two kinds of violence, one in which “perpetrators do not necessarily desire to harm victims, but they knowingly do so in order to achieve some other objective, such as selfish gain or even a greater moral good,” and a second kind of violence in which “victims are harmed because they see them as morally deserving of it.” Participants were then asked to rate the extent to which drone strikes framed around deaths to ISIS terrorists and collateral damage to Iraqi civilians were similar to each of the two kinds of violence on a five-point likert scale ranging from 1 (not similar at all) to 5 (completely similar). The two types of violence were not labeled as instrumental or moral. Items were counterbalanced for order.

Experiment 3 used a between-subjects design. After reading the vignettes (*SI Appendix*), participants rated their willingness to “physically break the man’s

thumb” for either the instrumental or moral reason provided on a seven-point likert scale ranging from 1 (not willing at all) to 7 (completely willing).

Experiments 4 and 5 used between-subjects designs. Participants in the instrumental conditions were asked to imagine a typical stranger they might encounter in everyday life (*SI Appendix*). Participants in the moral conditions were asked to further imagine that the person is “evil and enjoys making other people suffer.” Participants in the experimental conditions were asked to imagine pricking the stranger’s index finger with a pin, either in exchange for \$10 or as punishment for their immorality. After reading the vignettes, participants rated the stranger’s humanity by using similar likert scales to those used in Experiments 1 and 2 (*SI Appendix*).

ACKNOWLEDGMENTS. Funding for this research was provided by NSF Grant 1520031.

- Haslam N, Loughnan S (2014) Dehumanization and infrahumanization. *Annu Rev Psychol* 65:399–423.
- Haslam N (2006) Dehumanization: An integrative review. *Pers Soc Psychol Rev* 10: 252–264.
- Staub E (1989) *The Roots of Evil: The Origins of Genocide and Other Group Violence* (Cambridge Univ Press, Cambridge, UK).
- Smith D (2011) *Less Than Human: Why We Demean, Enslave, and Exterminate Others* (Macmillan, New York).
- Memmi A (2013) *The Colonizer and the Colonized* (Routledge, London).
- Davis D (2006) *Inhuman Bondage: The Rise and Fall of Slavery in the New World* (Oxford Univ Press, Oxford).
- Nussbaum M (1999) *Sex and Social Justice* (Oxford Univ Press, Oxford).
- Tedeschi J, Felson R (1994) *Violence, Aggression, and Coercive Actions* (Am Psychol Assoc, Washington, DC).
- Anderson CA, Bushman BJ (2002) Human aggression. *Annu Rev Psychol* 53:27–51.
- Manne K (2016) Humanism: A critique. *Soc Theory Pract* 42:389–415.
- Kelman H (1973) Violence without moral restraint: Reflections on the dehumanization of victims and victimizers. *J Soc Issues* 29:25–61.
- Bandura A (1999) Moral disengagement in the perpetration of inhumanities. *Pers Soc Psychol Rev* 3:193–209.
- Opatow S (1990) Moral exclusion and injustice: An introduction. *J Soc Issues Pol Rev* 46:1–20.
- Waytz A, Epley N, Cacioppo JT (2010) Social cognition unbound: Insights into anthropomorphism and dehumanization. *Curr Dir Psychol Sci* 19:58–62.
- Čehajić S, Brown R, González R (2009) What do I care? Perceived ingroup responsibility and dehumanization as predictors of empathy felt for the victim group. *Group Process Interg* 12:715–729.
- Rai TS, Fiske AP (2011) Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychol Rev* 118:57–75.
- Fiske AP, Rai TS (2014) *Virtuous Violence: Hurting and Killing to Create, Sustain, End, and Honor Social Relationships* (Cambridge Univ Press, Cambridge, UK).
- Pinker S (2011) *The Better Angels of Our Nature: Why Violence Has Declined* (Penguin, New York).
- Atran S (2010) *Talking to the Enemy: Faith, Brotherhood, and the (Un) making of Terrorists* (Harper Collins, New York).
- Boyd R, Gintis H, Bowles S, Richerson PJ (2003) The evolution of altruistic punishment. *Proc Natl Acad Sci USA* 100:3531–3535.
- Black D (1983) Crime as social control. *Am Sociol Rev* 48:34–45.
- Katz J (2008) *Seductions of Crime: Moral and Sensual Attractions in Doing Evil* (Basic Books).
- Ramsay G (2017) Dehumanization in religious and sectarian violence: The case of Islamic State. *Glob Discourse* 6:561–578.
- Appiah A (2008) *Experiments in Ethics* (Harvard Univ Press, Cambridge, MA).
- Gopnik A (June 5, 2006) Headless horseman: The reign of terror revisited. *The New Yorker*, p 80.
- Lang J (2010) Questioning dehumanization: Intersubjective dimensions of violence in the Nazi concentration and death camps. *Holocaust Genocide Stud* 24:225–246.
- Duff R (1990) *Intention, Agency and Criminal Liability: Philosophy of action and the criminal law* (Blackwell, Hoboken, NJ).
- Gray K, Wegner DM (2009) Moral typecasting: Divergent perceptions of moral agents and moral patients. *J Pers Soc Psychol* 96:505–520.
- Funk F, McGeer V, Gollwitzer M (2014) Get the message: Punishment is satisfying if the transgressor responds to its communicative intent. *Pers Soc Psychol Bull* 40:986–997.
- Miller E (1990) Executing minors and the mentally retarded: The retribution and deterrence rationales. *Rutgers Law Rev* 43:15–52.
- Bandura A, Barbaranelli C, Caprara G, Pastorelli C (1996) Mechanisms of moral disengagement in the exercise of moral agency. *J Pers Soc Psychol* 71:364–374.
- Waytz A, Epley N (2012) Social connection enables dehumanization. *J Exp Soc Psychol* 48:70–76.
- Kteily N, Bruneau E, Waytz A, Cotterill S (2015) The ascent of man: Theoretical and empirical evidence for blatant dehumanization. *J Pers Soc Psychol* 109:901–931.
- Bastian B, Laham SM, Wilson S, Haslam N, Koval P (2011) Blaming, praising, and protecting our humanity: The implications of everyday dehumanization for judgments of moral status. *Br J Soc Psychol* 50:469–483.
- Bastian B, Denson TF, Haslam N (2013) The roles of dehumanization and moral outrage in retributive justice. *PLoS One* 8:e61842.
- Maoz I, McCauley C (2008) Threat, dehumanization, and support for retaliatory aggressive policies in asymmetric conflict. *J Conflict Resolut* 52:93–116.
- Gray K, Young L, Waytz A (2012) Mind perception is the essence of morality. *Psychol Inq* 23:101–124.
- Fiske ST (2009) From dehumanization and objectification to rehumanization: Neuroimaging studies on the building blocks of empathy. *Ann N Y Acad Sci* 1167:31–34.
- Harris LT, Fiske ST (2006) Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychol Sci* 17:847–853.
- Harris L, Fiske S (2011) Perceiving humanity or not: A social neuroscience approach to dehumanized perception. *Social Neuroscience: Toward Understanding the Underpinnings of the Social Mind*, eds Todorov A, Fiske S (Oxford Univ Press, Oxford), pp 123–134.
- Appiah K (2003) *Thinking it Through: An Introduction to Contemporary Philosophy* (Oxford Univ Press, Oxford).
- Gray HM, Gray K, Wegner DM (2007) Dimensions of mind perception. *Science* 315: 619.
- Leyens J, Rodriguez-Torres A, Rodriguez-Perez A, Gaunt R, Paladino M (2001) Psychological essentialism and the differential attribution of uniquely human emotions to ingroups and outgroups. *Eur J Soc Psychol* 31:395–411.
- Schwarz N (1999) Self-reports: How the questions shape the answers. *Am Psychol* 54: 93–105.
- Smith D (2016) Paradoxes of dehumanization. *Soc Theory Pract* 42:416–443.
- Sinott-Armstrong W, Wheatley T (2014) Are moral judgments unified? *Philos Psychol* 27:451–474.
- Cuddy AJ, Fiske ST, Glick P (2007) The BIAS map: Behaviors from intergroup affect and stereotypes. *J Pers Soc Psychol* 92:631–648.
- Haslam N, Bain P, Douge L, Lee M, Bastian B (2005) More human than you: Attributing humanness to self and others. *J Pers Soc Psychol* 89:937–950.
- Khamitov M, Rotman JD, Piazza J (2016) Perceiving the agency of harmful agents: A test of dehumanization versus moral typecasting accounts. *Cognition* 146:33–47.
- Ward AF, Olsen AS, Wegner DM (2013) The harm-made mind: Observing victimization augments attribution of minds to vegetative patients, robots, and the dead. *Psychol Sci* 24:1437–1445.
- Rudman LA, Mescher K (2012) Of animals and objects: Men’s implicit dehumanization of women and likelihood of sexual aggression. *Pers Soc Psychol Bull* 38:734–746.
- Viki G, Osgood D, Phillips S (2013) Dehumanization and self-reported proclivity to torture prisoners of war. *J Exp Soc Psychol* 49:325–328.
- Leidner B, Castano E, Ginges J (2013) Dehumanization, retributive and restorative justice, and aggressive versus diplomatic intergroup conflict resolution strategies. *Pers Soc Psychol Bull* 39:181–192.
- Goff PA, Eberhardt JL, Williams MJ, Jackson MC (2008) Not yet human: Implicit knowledge, historical dehumanization, and contemporary consequences. *J Pers Soc Psychol* 94:292–306.
- Margalit A (2002) *The Ethics of Memory* (Harvard Univ Press, Cambridge, MA).
- Schein C, Gray K (2015) The unifying moral dyad: Liberals and conservatives share the same harm-based moral template. *Pers Soc Psychol Bull* 41:1147–1163.
- Turiel E (1983) *The Development of Social Knowledge: Morality and Convention* (Cambridge Univ Press, Cambridge, UK).
- Mikhail J (2007) Universal moral grammar: Theory, evidence and the future. *Trends Cogn Sci* 11:143–152.
- Bushman BJ, Anderson CA (2001) Is it time to pull the plug on the hostile versus instrumental aggression dichotomy? *Psychol Rev* 108:273–279.
- Cuddy A, Rock M, Norton M (2007) Aid in the aftermath of Hurricane Katrina: Inferences of secondary emotions and intergroup helping. *Group Process Interg* 10:107–118.
- Cikara M, Farnsworth RA, Harris LT, Fiske ST (2010) On the wrong side of the trolley track: Neural correlates of relative social valuation. *Soc Cogn Affect Neurosci* 5:404–413.
- Tam T, et al. (2007) The impact of intergroup emotions on forgiveness in Northern Ireland. *Group Process Interg* 10:119–136.
- Buhrmester M, Kwang T, Gosling SD (2011) Amazon’s mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspect Psychol Sci* 6:3–5.