Denying Humanity: The Distinct Neural Correlates of Blatant Dehumanization

Emile Bruneau
Massachusetts Institute of Technology and
University of Pennsylvania

Nir Jacoby
Massachusetts Institute of Technology and
Columbia University

Nour Kteily
Northwestern University

Rebecca Saxe
Massachusetts Institute of Technology

Recent behavioral work demonstrates that many people view low-status groups as less "evolved and civilized" than high-status groups. Are these people using blatant expressions of dehumanization simply to express strong dislike toward other groups? Or is blatant dehumanization a process distinct from other negative assessments? We tested these competing hypotheses using functional neuroimaging. Participants judged 10 groups (e.g., Europeans, Muslims, rats) on four scales: blatant dehumanization, dislike, dissimilarity and perceived within-group homogeneity. Consistent with expectations, neural responses when making ratings of dehumanization diverged from those when judging the same targets on the other related dimensions. Specifically, we found regions in the left inferior parietal cortex (IPC) and left inferior frontal cortex (IFC) that were selectively parametrically modulated by dehumanization ratings. The pattern of responses in the left IFC was also consistent with animalistic dehumanization: high responses to low-status human groups and animals, and lower responses to high-status human groups. By contrast, a region in the posterior cingulate cortex was parametrically sensitive specifically to liking. We therefore demonstrate a double dissociation between brain activity associated with judgments of blatant dehumanization and judgments of dislike.

Keywords: blatant dehumanization, dehumanization, prejudice, fMRI, social cognition

Supplemental materials: http://dx.doi.org/10.1037/xge0000417.supp

The depiction of some groups as less than fully human is an endemic feature of intergroup relations. In particular, low-status groups (e.g., Native Americans, Roma/gypsies, Africans) have routinely been portrayed throughout history as animals, casting them outside of the purview of moral concern (Bandura, Underwood, & Fromson, 1975; Kelman, 1973) and making them more susceptible to exploitation through slavery, colonization, and genocide (Smith, 2011). Even favored members of these groups have often been perceived as savages in need of "civilizing," resulting in paternalistic policies such as the forced removal of native children to be educated in Western boarding schools (Haslam, 2006; Jahoda, 2014).

Despite its central importance to intergroup hostility and human history, however, relatively little is known about the neural processes that underlie dehumanization, or that distinguish it from related processes. In particular, questions remain about the degree of overlap between dehumanization and dislike (Haslam & Loughnan, 2014). The current research aimed to examine the neural activity associated with making blatant dehumanization judgments. In particular, we sought to determine if the neural processes associated with consciously and actively dehumanizing another group are similar or distinct from the neural processes associated with consciously judging the likability of another group. Based on theoretical work, which we discuss further below, we hypothesized that judging the humanity versus likability of groups would be associated with activity in distinct neural populations. As an exploratory extension, we also sought to determine how the neural processes associated with dehumanization and dislike judgments related to judgments on two other conceptually related but plausibly distinct dimensions: the perceived homogeneity of other groups and their perceived dissimilarity to the self.

Blatant Dehumanization

Much modern empirical research on dehumanization has been predicated on the assumption that open depictions of others as less than fully human is largely a thing of the past, and that dehumanization remains today in our general assessments about outgroup others not as a blatant expression, but as a subtle or implicit judgment (see Kteily & Bruneau, 2017b). To identify subtle forms of "everyday" dehumanization researchers have developed indirect and covert measures. For example, people selectively attribute fewer complex, uniquely human emotions (e.g., guilt, pride) or traits (e.g., rude, ambitious) to members of certain out-groups than they do to members of their own group (Haslam, 2006; Leyens et al., 2000). The subtlety of these measures comes from the fact that participants are not explicitly told that they are being asked to make judgments of humanity—indeed, discerning participants may recognize that the emotions/traits are positively or negatively valenced, but are very unlikely to identify the human-specificity of these emotions and traits as a relevant dimension of interest. Of note, subtle dehumanization may occur relatively passively, when people overlook opportunities to spontaneously consider another person's mind (Epley & Waytz, 2010; Harris & Fiske, 2006; Waytz & Schroeder, 2014). Indeed, as Haslam and Loughnan (2014) note, subtle dehumanization can reflect an unintentional lack of association between a target and humanness (e.g., Martinez et al., 2012), in addition to an unconscious positive association between the target and animals (e.g., Goff, Eberhardt, Williams, & Jackson, 2008).

Although the focus on subtle dehumanization has been instructive and important in shaping our understanding of intergroup perceptions, recent research has demonstrated that blatant expressions of dehumanization—in which individuals explicitly rate another target as more animal-like—are not relegated to a colonial past, but remain pronounced features of many intergroup contexts (see Kteily & Bruneau, 2017b, for a review). For example, using a measure of dehumanization based on the popular "Ascent of Man" diagram, Kteily, Bruneau, Waytz, and Cotterill (2015) showed that contemporary Americans and Europeans openly dehumanize low-status groups (e.g., Muslims) by rating them further than their own groups from "fully evolved" humans (see also Linden, Bjorklund, & Backstrom, 2016; Pacilli, Roccato, Pagliaro, & Russo, 2016). As noted in Kteily and Bruneau (2017b), it is probable that people generally use the Ascent Dehumanization scale metaphorically to express the view that some targets are "superior" to others on dimensions that clearly differentiate animals from humans (such as morality, "civilization," and mental capacities), rather than using it to express a literal view of targets as genetically more similar to nonhumans. This form of blatant and deliberative dehumanization is highly consequential: Dehumanization on the "Ascent" scale predicts aggressive outcomes, such as support for torture and reluctance to provide aid to outgroup civilian victims of violence (Kteily et al., 2015), support for armed conflict with Iran (Kteily, Hodson, & Bruneau, 2016), and support for hostile policies against Muslim refugees and Mexican immigrants (Bruneau, Kteily, & Laustsen, 2017; Kteily & Bruneau, 2017a). Empirically, blatant and subtle dehumanization are only weakly correlated (rs ~ .20; Kteily et al., 2015), and when compared directly, blatant dehumanization more strongly predicts intergroup behaviors (Bruneau et al., 2017; Kteily et al., 2015).

Blatant Dehumanization and Dislike

Blatant dehumanization is therefore a pervasive and consequential feature of intergroup relations in the real world. However, one important concern with blatant measures of dehumanization is that people may be using them simply as a convenient way to express strong dislike, making them "a manifestation of antipathy" (Haslam & Loughnan, 2014). Evidence suggesting that blatant dehumanization and dislike scales could capture a common psychological process comes from previous research showing that ratings of ascent dehumanization and ratings on a feeling thermometer are consistently and robustly correlated with each other, across a range of cultural contexts (rs ~ .5–.6; Kteily & Bruneau, 2017a; Kteily et al., 2015, 2016). The question therefore remains, whether blatant measures of dehumanization are capturing the denial of others' humanity, or if they merely reflect a common (if extreme) version of "othering" we sometimes engage in when construing those who do not belong to our group as unlikeable (or dissimilar or homogeneous, as we discuss below).

Despite the high correlations between ratings of blatant dehumanization and dislike, we propose that dehumanization and dislike are meaningfully distinct constructs. Dehumanization does not always imply dislike. We can consciously deny a full human mind to children and dogs by attributing them less agency (cf. Gray, Gray, & Wegner, 2007), even though we have great fondness toward them. Similarly, the depiction of indigenous people as "noble savages" combines dehumanization with paternalistic warmth. We can also humanize those we dislike, for example by recognizing the fully human capacities of a brilliant but arrogant colleague. Even when blatant dehumanization and dislike are both present, blatant dehumanization predicts hostile attitudes and behaviors toward low-status outgroups when controlling for liking or warmth (e.g., feeling thermometer ratings or racial resentment: Jardina & Piston, 2016; Kteily et al., 2015, 2016), suggesting that these measures may in fact capture distinct psychological processes. In intergroup conflict, blatant dehumanization might be particularly associated with the view of the outgroup as morally inferior and deserving of (or even *necessitating*) expulsion or extermination (Fiske & Rai, 2014).

In sum, based on behavioral data, the question of whether the processes underlying judgments of blatant dehumanization and dislike are distinct remains the subject of important debate. In the research reported here, we sought to examine the distinction between blatant dehumanization and dislike using a methodological alternative to behavioral data: functional MRI (fMRI). Self-report and behavioral measures provide one reasonable way to examine the relationship between dehumanization and dislike. However, judgments that appear behaviorally distinct could nevertheless share considerable overlap in the cognitive systems and neural machinery that they recruit when making the judgments. For example, behavioral ratings of dislike and hatred may correlate as strongly with each other as do dislike and dehumanization, but these pairs of processes may have different degrees of overlap in the neural responses they recruit. That is, the degree of similarity in self-report ratings cannot definitively determine the neural overlap or cognitive similarity of the processes. fMRI is well-suited to examine this issue, because it can determine whether similar or distinct brain regions are recruited when making overt judgments of blatant dehumanization and dislike, an important question that

cannot be ascertained from the moderate to strong correlations observed between these constructs at the behavioral level.

Dehumanization, Dislike, Dissimilarity, and Homogeneity

Given continued debate about whether blatant measures of dehumanization are distinguishable from antipathy, our central interest in the current research was to compare neural responses when making deliberative judgments of dehumanization versus dislike. On the theoretical grounds noted above, our a priori hypothesis was that the brain regions activated while making these judgments would be distinct.

At the same time, we took the opportunity to expand our investigation beyond dislike, examining the overlap in neural responses when making judgments of dehumanization versus judgments on two other forms of group evaluation: dissimilarity from the self, and perceived homogeneity of group members. Although not previously compared directly with ascent dehumanization (or other measures of blatant dehumanization), homogeneity and dissimilarity are conceptually related to dehumanization. For example, classic theorizing about dehumanization suggests that one hallmark of the phenomenon is the tendency to see another group as a homogenous "undifferentiated mass" (Kelman, 1973). Additionally, previous empirical work suggests that individuals tend to withhold full humanity for the self and ingroup members relative to (more distant) others (Haslam, 2006; Kteily et al., 2015; Leyens et al., 2000). It is therefore plausible that the degree of dehumanization of others simply reflects the extent to which they are judged to be homogeneous and/or different from the self (and the ingroup), which could be reflected in overlapping neural activity when engaged in dehumanization and similarity/homogeneity judgments. On the other hand, individuals could dehumanize outgroups even if they perceived some hetereogeneity among them (perhaps subtyping those they humanized as exceptions that prove the rule), and sometimes even targets similar to us in important ways can be dehumanized (e.g., reciprocal dehumanization of Israelis and Palestinians of one another despite many cultural similarities; see Bruneau & Kteily, 2017).

Psychologically, one factor that may well distinguish humanity attribution from judgments of dislike, homogeneity, and dissimilarity is the extent to which it is tied to considerations of social status. That is, we tend to dehumanize those we see as *beneath* us—falling short on key dimensions (e.g., intelligence, "civilization," morality) that elevate humans above lower animals, which are considered inherently "inferior." Consistent with the centrality of social status to humanity attribution, previous behavioral research has shown that individuals most concerned with maintaining hierarchical differentiation between groups (i.e., those highest on social dominance orientation; Ho et al., 2015) are also those most likely to blatantly dehumanize low-status groups (Kteily et al., 2015). Although social rank is unlikely to be entirely irrelevant to judgments of likability, similarity, or homogeneity, it may be less central to these judgments than it is to judgments of dehumanization: We regularly judge targets to be our equals (or superiors), even though we dislike them (e.g., an annoying boss), find them dissimilar (e.g., members of culturally dissimilar but respected countries) or think of them as homogenous (e.g., Americans struggling to differentiate between individual Japanese, even

if they are enthralled by Japanese technological and cultural achievements).

Evidence that the neural processes associated with dehumanization judgments are distinguishable from those underlying judgments of dissimilarity and homogeneity (in addition to those underlying judgments of dislike) would enhance confidence that dehumanization is a distinguishable cognitive process (despite its potential to be associated with perceptions of similarity and/or homogeneity).

Previous Neuroimaging Research

No past research has examined neural activity when people are actively judging targets' humanity, or tried to distinguish the neural responses active during blatant dehumanization judgments versus when making conceptually similar judgments. However, a few prior studies have examined spontaneous neural activity when *passively* viewing images of typically dehumanized targets (e.g., Harris & Fiske, 2006, 2007; Krendl, 2016), images of animals and humans, or images depicting animal-like and human-like behavior (Jack, Dawson, & Norr, 2013).

In a first line of research, participants in a neuroimaging study were presented with images of people from marginalized groups perceived to be low in warmth and competence (homeless, drug addicts with drug paraphernalia) and images of people from other groups higher in warmth and/or competence (e.g., college students, fire fighters in firefighting gear). Across two studies, passively viewing images of low warmth/competence group members was associated with less activity in the MPFC than viewing images of people from groups higher in warmth/competence (Harris & Fiske, 2006, 2007). Because the MPFC has been implicated in mentalizing (Saxe, 2006; Saxe & Powell, 2006), the authors interpreted their findings to reflect the dehumanization of low-warmth/low-competence targets by denying them mind. Notably, the targets in these studies were not overtly dehumanized in the stimuli, and Harris and Fiske (2006, 2007) therefore interpret the observed activity to reflect a subtle form of dehumanization that may even be beyond conscious awareness. This result was replicated by an independent researcher, who also examined neural responses to passively viewing similar groups (e.g., homeless people vs. businessmen; Krendl, 2016). At the same time, the MPFC has also been implicated in self/other judgments (Mitchell, Macrae, & Banaji, 2006), and it is therefore possible that the reduced activation in MPFC when encountering low-warmth/low-competence targets was rooted in an apprehension of these targets' dissimilarity from the self, rather than a tendency to overlook their humanity.

In another line of research, human/animal distinctions were examined in two studies that depicted targets in more overtly dehumanizing ways: The first contrasted neural activity when participants viewed images of static human faces versus images of animals in motion (e.g., a bat in flight, a prowling leopard), and the second examined neural activity when people were provided with images of humans paired with descriptions that clearly evoked animalistic associations (e.g., drinking from a puddle, defecating in public) versus images of humans paired with particularly humanizing associations (e.g., engaging in moral behavior, taking an exam; Jack et al., 2013). Compared with viewing images of humans and humanizing behaviors, viewing images of animals and animalistic behaviors consistently resulted in greater activity in left

lateralized brain regions, including the inferior frontal cortex (IFC). Notably, the images in the Jack et al. (2013) experiments contrast with those used by Harris and Fiske (2006, 2007) in the extent to which they depict the targets in ways that explicitly liken them to animals. As such, the neural activity observed by Jack et al. (2013) may reflect spontaneous but still relatively blatant dehumanization.

By providing participants with rich visual stimuli in the absence of judgments about the target images/groups, the studies reviewed above have the advantage of assessing individuals' spontaneous responses. However, one downside of this approach is that it makes it difficult to determine whether activation patterns reflect dehumanization per se, or some other related process across which the conditions differ. For example, when participants are presented with complex stimuli such as images of humans in social context, it is difficult to determine if the neural activity generated when viewing images of low- versus high-status humans (e.g., college student vs. drug addict), or humans versus animals (e.g., human face vs. bat) are associated with differences between the groups on dimensions other than their "human-ness," such as the context they are in (college dorm vs. living on the streets), or the familiarity, likability or perceived similarity of the targets to the self. It is therefore unclear if activity is associated with a *perceptual* distinction (e.g., static vs. biological motion, familiar vs. unfamiliar social environment) or a *conceptual* distinction; and if conceptual, whether the activation is unique to dehumanization, per se, or a range of related (but unmeasured) dimensions.

The approach applied here directly examines neural responses to deliberative dehumanization judgments about groups, and contrasts this to judgments on related dimensions. One disadvantage of our approach is that by explicitly asking individuals to make judgments about targets' humanity, likability, similarity to the self, and homogeneity, our approach may be less close than prior efforts to capturing the ways in which individuals naturalistically interact with targets in society. That said, active deliberation about a target's human capacities, intentions, and morals may be critical to a range of consequential intergroup decisions, such as determining culpability and punishment in legal settings, rejection of intergroup treaties, opposition to immigration, support of torture, and so forth. It is also the case that some of the activation we observe may have to do with features of the rating scales we use, or cognitive processes associated with self-presentational concerns that some participants might experience when making their judgments. At the same time, an important advantage of our approach is that, by contrasting activation to dehumanization judgments with judgments on potentially overlapping dimensions, it provides greater confidence that any neural activity associated with dehumanization judgments is indeed reflective of humanity attribution per se. Moreover, and critically, to the extent that regions identified as unique to dehumanization using our approach converge with those identified by prior researchers to potentially be involved in spontaneous dehumanization while passively engaging with naturalistic stimuli, we can be particularly confident in these regions' importance (and distinctiveness) to dehumanization.

The current research joins a handful of other studies that have examined neural responses to deliberative social judgments. For example, Farrow et al. (2011) examined neural responses while making explicit judgments about image pairs, including judgments about social status. In this study, participants were presented with

a pair of social targets and asked to rate which one was higher or lower on a specific dimension: age, gender, fame, and social status. As with the current research, the benefit of this approach is that it examines unique activation associated with different judgments about the same targets, to identify activation uniquely associated with dimensions of key theoretical interest. The authors found that making social hierarchy judgments versus all other judgments was associated with greater activity in visual cortical, fusiform face area (FFA), and a region in left inferior frontal cortex (IFC). The left IFC activity is notable in that it overlaps with the left IFC region active when viewing animals versus humans, and humans engaging in animal-like versus human-like behavior (Jack et al., 2013). This may have to do with the fact that animals are perceived not only as *different* than humans but also as *beneath* them in the social order (e.g., Haslam & Loughnan, 2014; Hodson, Kteily, & Hoffarth, 2014; Smith, 2011).[1] As noted above, perceptions of a group as inferior may be especially important to denials of humanity relative to other conceptually related attributions (e.g., Smith, 2011).

Current Research

Here, we directly examined neural responses to dehumanization judgments about groups (including high- and low-status human groups and animals), in comparison with judgments on related dimensions, by explicitly asking individuals to make judgments about targets' humanity, likability, similarity to the self, and homogeneity. In the current study our primary goal was to test the hypothesis that judgments of blatant dehumanization (assessed using the ascent measure of blatant dehumanization) are neurally distinct from judgments of dislike (assessed using feeling thermometers). We also predicted that judgments of blatant dehumanization would be neurally distinct from judgments of dissimilarity and homogeneity.

To test these hypotheses, we measured BOLD signal while participants judged a common set of target groups using the blatant dehumanization, dislike, dissimilarity and homogeneity scales during functional MRI (fMRI). That is, we sought to identify brain regions differentially associated with explicitly evaluating the same target along distinct dimensions (i.e., an effect of task, independent of the stimulus, similar to Farrow et al., 2011, and others: Harel, Kravitz, & Baker, 2014; Spunt, Falk, & Lieberman, 2010; Spunt, Satpute, & Lieberman, 2011).

The groups assessed in the present study included high-status human groups (Americans, Europeans, surgeons), low-status human groups (homeless, Muslims, Gypsies)[2] and animals (rats, puppies). We chose groups that would plausibly be rated at values spanning both the dehumanization and dislike scales. In particular, by including both a disliked animal (rats) and a liked animal (puppies), we hoped to include targets that might be expected to most clearly separate dehumanization from dislike.

A second goal was to compare neural responses while making dehumanization (and dislike) judgments to judgments on two other

---

[1] However, we acknowledge that the activation observed in Farrow et al. (2011) may be due to something other than status judgments, such as complexity of the judgments being rendered.

[2] An independent sample of participants confirmed these groups to be high and low in status—see online supplementary study.

scales: dissimilarity from the self and perceived within-group homogeneity. We sought to determine both if brain regions were sensitive to dehumanization judgments *controlling* for judgments on the other scales, and—more conservative still—if those regions were significantly *more* sensitive to dehumanization judgments than judgments on the other scales.

A final goal was to determine mean responses to specific groups within any brain regions that were selectively parametrically sensitive to blatant dehumanization judgments. Because previous work has shown that low-status groups are often animalistically dehumanized (e.g., Goff et al., 2008; Kteily et al., 2015; Lammers & Stapel, 2011), we examined if regions that show a selectively sensitivity to judgments of humanity (vs. related judgments) showed different mean levels of activation when judging high-status human groups versus low-status human groups and animals (consistent with work suggesting that the tendency to dehumanize is associated with attunement to hierarchical differentiation between high- and low-status groups; e.g., Kteily et al., 2015).

In sum, we reasoned that the most compelling evidence for activation uniquely subserving dehumanization would be activation in a given region that (a) was significantly parametrically modulated by dehumanization judgments, (b) remained significantly associated with dehumanization judgments controlling for potentially overlapping judgments, (c) was significantly *more strongly* associated with dehumanization judgments than any other judgments, (d) responded differently to judgments of high-status groups as compared with low-status groups and animals, and (e) converged with activation observed in dehumanization studies using different methods.

Method

This research was approved by the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects (COUHES).

Participants

We collected data from as many participants as possible, within the time constraints of the funding source. Twenty-seven adults were recruited for the study from the greater Boston area. Data from three participants were excluded due to movement (>90 total events with activity >3 standard deviations from the mean and 2 mm composite movement between consecutive volumes), leaving 24 adults (M_{age} = 27.8, SD = 7.4; 11 female).[3] Participants had no history of neurological or psychiatric impairment, and had normal or corrected vision. All participants provided informed consent. No further exclusion criteria were used.

Paradigm

Participants were recruited for a study on "social categorization," and were told prior to scanning that they would be rating 10 social groups or animate categories (Americans, Europeans, Muslims, Gypsies, homeless, toddlers, surgeons, robots, rats, puppies) on four different metaphorical scales: blatant dehumanization (as captured by "evolvedness" ratings on the Ascent Dehumanization scale; Kteily & Bruneau, 2017a; Kteily et al., 2015, 2016) liking (the feeling thermometer; Haddock, Zanna, & Esses, 1993); sim-

ilarity (using the inclusion of outgroup in the self scale; Schubert & Otten, 2002), and heterogeneity (using a scale developed for the purposes of this study and conceptually based on Park & Judd, 1990). Each scale used five images to indicate a range of responses on the psychological measure (e.g., thermometers depicting different levels of mercury from cold to hot to represent liking), with the images progressing in favorability from left to right (see Figure 1). Because most people have a strong spatial association between the right side of space and positive concepts (Willems, Hagoort, & Casasanto, 2010), we kept a consistent spatial mapping of all scales across trials and participants. Ratings were made on a 21-point continuous scale (i.e., 0–20). A list of the groups and an image of all the scales were provided prior to scanning to familiarize the participants with the task and each of the measurement tools. No participants expressed any confusion with the metaphorical depictions.

For each trial, the name of one of the groups (e.g., "homeless") appeared above one of the scales (e.g., ascent dehumanization), and a rating arrow appeared below the scale at either the left or right anchor (randomized across trials and runs). Within each run, stimuli were presented in a jittered, event-related design. Participants were given 6 s to move the arrow left or right using two buttons on a scanner safe button box—the final position indicated their rating. Every group was rated on every scale exactly once in each run, for a total of 40 trials/run; individual trials were separated by 39 jittered interstimulus intervals of 4 s ($\times$ 10), 6 s ($\times$ 19), or 8 s (10 $\times$) (see Figure 2). A rest period of 12 s was added to the beginning and end of the run to help facilitate calculation of baseline activity, resulting in a total run time of 502 s.

Participants engaged in six to eight runs, constrained by time and self-reported fatigue (four completed six runs, two completed seven runs, 18 completed all eight runs). Order of stimuli was randomized across runs, and participants were specifically instructed to avoid trying to remember what they had answered in previous runs, and to instead treat each rating in Runs 2–8 as if it was the first time they were making the rating, so that they were using a similar cognitive process each run (i.e., social categorization rather than recall).[4]

Data Acquisition

Participants were scanned using a Siemens Magnetom Tim Trio 3T system (Siemens Solutions, Erlangen, Germany) in the Athinoula A. Martinos imaging center at the McGovern Institute for Brain Research at MIT using a 32-channel head coil. Functional images were acquired with near whole brain coverage, in 32 near axial 64 $\times$ 64 slices (voxel size: 3.125 mm $\times$ 3.125 mm $\times$ 3.13 mm; 0.313 mm interslice spacing, TR = 2s, TE = 30 ms, flip angle = 90). High-resolution structural (anatomical) images were acquired using a T1 MPRAGE sequence (voxel size: 1 mm $\times$ 1 mm $\times$ 1 mm).

---

[3] We did not collect any demographic information beyond age and gender.

[4] Across all identified ROIs we found that beta values in Run 1 were just as strongly correlated with beta values in subsequent runs as were beta values in Runs 2–8 with each other, suggesting that participants were not using markedly different neural processes in Run 1 versus Runs 2–8.
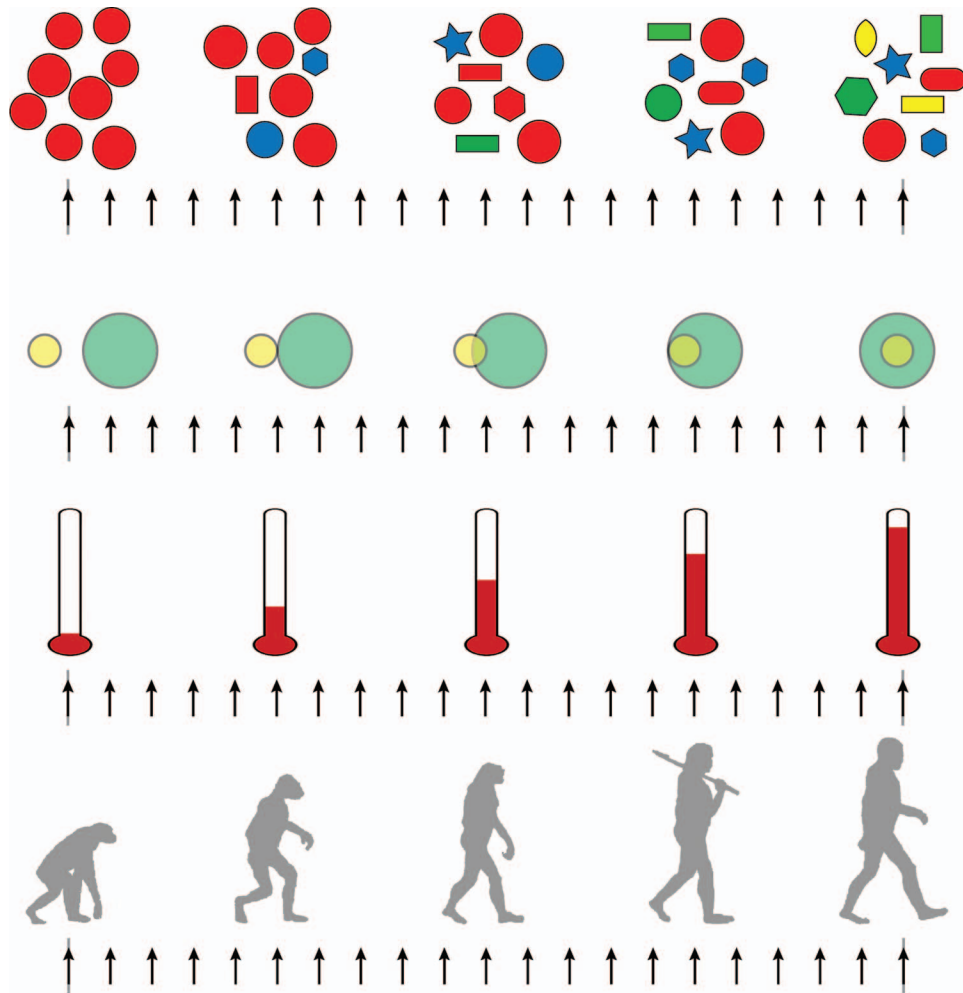
Figure 1. Scale images for homogeneity–heterogeneity, dissimilarity–similarity, dislike–like, dehumaniza-tion–humanization. Arrows indicate all possible rating positions. For each trial, a group appeared above one of the four scales, and a single arrow appeared randomly at the left or right anchor position. Over the next 6 s the participant moved the arrow using left and right buttons, until the arrow reached the position that indicated the participant's rating for that group on that scale. Each of 10 groups (Americans, Europeans, surgeons, robots, toddlers, Muslims, Gypsies, homeless, puppies, rats) was rated on each scale. Participants saw each combination of group and scale in randomized order. See the online article for the color version of this figure.

Data Preprocessing and Modeling

Data were processed using SPM8 and SnPM, supplemented by custom MATLAB scripts. Functional data was motion corrected using rigid body transformation. Data from each run was registered to the first volume of the first functional run. The functional runs were coregistered to the anatomical scan, which was coregistered and normalized onto a common brain space (Montreal Neurological Institute, EPI Template). Data were smoothed using a Gaussian filter (full width half maximum = 5 mm).

Data Analysis

A primary goal of this study was to determine if neural processes recruited when making blatant dehumanization judgments are distinct from those recruited when making three conceptually related judgments: dissimilarity, homogeneity, and (in particular) dislike. Whole brain analyses were used to identify regions where univariate responses were sensitive to parametric ratings of dehumanization, dislike, dissimilarity, or homogeneity. In principle, brain regions could display sensitivity to dehumanization if (a) mean responses across groups were stronger while making dehumanization judgments versus judgments of the same groups on the other scales, and/or (b) activity in a brain region was correlated with behavioral ratings of dehumanization, but not with behavioral ratings on the other scales. The stimuli in the current study were better matched when making ratings on the same scale than they were across scales, particularly because the silhouettes of human/humanoid forms on the dehumanization scale may themselves generate activity in a brain regions associated with social cognition (e.g., fusiform face area [FFA]; Kanwisher, McDermott, & Chun,
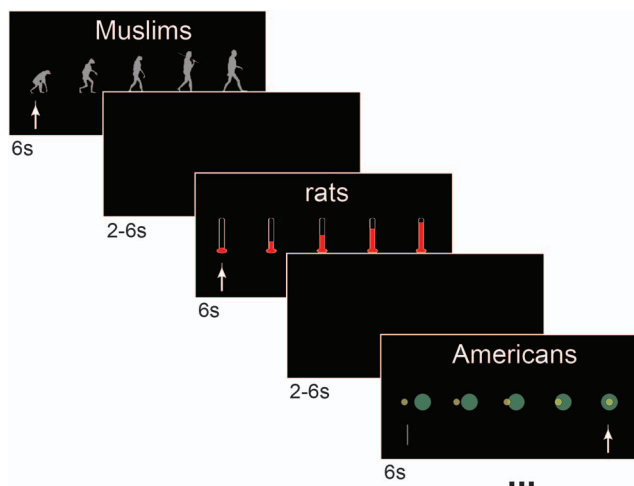
Figure 2. Experimental setup. For each trial, a group appeared above one of the four scales: (de)humanization, (dis)like, (dis)similarity, (homo)heterogeneity; rating arrow appeared randomly at the left or right anchor position. Over the next 6 s the participant moved the arrow using left and right buttons to make their rating on the continuous scale; each rating was followed by a 2 s–6 s jittered interval. In each run, each of 10 groups (Americans, Europeans, surgeons, robots, toddlers, Muslims, Gypsies, homeless, puppies, rats) was rated on each scale (i.e., 40 ratings per run). Participants saw each combination of group and scale in randomized order. See the online article for the color version of this figure.

1997; extrastriate body area [EBA]; Downing, Jiang, Shuman, & Kanwisher, 2001), relative to the geometric shapes of the other scales. The study was therefore best suited to identify brain regions that displayed selective parametric sensitivity within a rating scale. Parametric analysis is a common fMRI analysis strategy that has been used extensively to provide information about a relationship between neural activity and a stimulus parameter (e.g., word presentation rate [Büchel, Holmes, Rees, & Friston, 1998], emotional valence [Heinzel et al., 2005], monetary reward [Knutson, Adams, Fong, & Hommer, 2001], behavioral response [e.g., reaction time [RT], Yarkoni, Barch, Gray, Conturo, & Braver, 2009], affective pain rating [Rainville, Duncan, Price, Carrier, & Bushnell, 1997], "feeling of knowing" [Kikyo, Ohki, & Miyashita, 2002]). We therefore used parametric analyses to define regions that were sensitive to judgments across the four scales.

We analyzed the neuroimaging data using whole-brain general linear models (GLMs). First, we asked whether the same, or different, neural regions were parametrically associated with behavioral ratings of blatant dehumanization, dislike, dissimilarity, and homogeneity. This model (Model 1) included, for each scale: an average scale regressor, and a (mean-centered) parametric regressor constructed from the in-scanner rating on each trial. Regressors were defined as boxcar functions covering the 6-s duration of the trial. All regressors were convolved with a canonical double-gamma hemodynamic response function. We examined both positive and negative contrasts for each regressor (e.g., brain regions correlated with increased humanization, vs. increased dehumanization, of the target groups). Results from these whole brain random effects analyses were corrected for multiple comparisons in SnPM (Nichols & Holmes, 2002), which performs

Monte Carlo permutation tests on the collected data to establish empirical null distributions for the peak T and cluster size in each analysis. Peak and cluster were equally weighted in the analysis, and the corrected threshold was set at $p < .05$; this process also avoids the documented deficiencies of the SPM default corrections for multiple comparisons (Eklund, Nichols, & Knutsson, 2016).

Second, we used the results from the first model to define regions of interest that could constrain our analysis with a second model. We did this by using the results from Model 1 to identify regions with a consistent parametric response to one task in n-1 subjects ($p < .001$, uncorrected; Esterman, Tamber-Rosenau, Chiu, & Yantis, 2010). In the remaining, left-out subject, we created a second model, including a regressor for each of the 40 unique items (10 targets $\times$ 4 scales; boxcar convolved with a standard HRF). We estimated the response in each region (defined by n-1 subjects, in Model 1) to each item (in the nth subject, in Model 2), thus ensuring that data contributing to the definition of a region was never extracted from that region (Kriegeskorte, Lindquist, Nichols, Poldrack, & Vul, 2010). This process was iterated across all subjects.

For each of the regions of interest (ROIs), we calculated the mean brain response beta for each of the evaluations (10 targets $\times$ 4 scales) per individual. In addition, we calculated the average behavioral score for each of the evaluations. Then, for every evaluation scale we built a multilevel model predicting the brain response while engaging in each type of evaluation from all of the behavioral scores. We also included in the model a random intercept per subject, which was used to account for within-subject variability. The models were done using "lme4" and "lmertest" packages for *R*, using the following formula: *beta(evaluation)* $\sim$ *dehumanization + dislike + dissimilarity + homogeneity + (1|subject)*. With this model we examined (a) if the region remained sensitive to the rating scale that was used to define it when controlling for ratings on all other scales (Satterthwaite's approximation using the calcSatterth function in lmerTest); and (b) if the correlation between neural responses and ratings was significantly stronger for the scale used to define the region versus each of the others (Malgady, 1987).

Finally, in a secondary analysis we tested whether mean activity in the brain regions that were parametrically sensitive to dehumanization judgments distinguished between high-status human groups, low-status human groups and animals in a way consistent with animalistic dehumanization. To do this, we used a 4 scale (dehumanization, dislike, dissimilarity, homogeneity) $\times$ 3 target group (animals, low-status humans, high-status humans) ANOVA, with target group as a within-subjects factor. We then performed planned paired-samples *t* tests to determine if the pattern of mean responses matched the most rigorous test for dehumanization-specific brain regions: mean responses that are higher for low-status groups and animals than high-status groups, specifically when making dehumanization judgments. Paired *t* tests were Bonferroni corrected for the three tests within each scale (high-status vs. low-status, high-status vs. animals, low-status vs. animals), and corrected *p* values are reported.

In sum, we performed parametric whole brain analysis to examine our primary research question: identifying where neural activity is modulated by blatant dehumanization, and especially whether these regions are distinct from regions implicated in judgments of liking, similarity or homogeneity. We then per-

formed ROI analyses to specifically query the regions parametrically associated with dehumanization. The ROI analyses allowed us to determine (a) whether parametric sensitivities were specific to one rating scale, controlling for the others; and (b) whether the dehumanization-sensitive brain regions distinguished between low- and high-status human groups (rather than just being driven by, e.g., the animal targets). Overall, these analyses allowed us to identify brain regions that were specifically sensitive to dehumanization judgments, and which showed a pattern of mean response consistent with the blatant dehumanization of low-status human groups.

Results

Behavioral Results

Participants rated each group on each scale consistently across runs (average split-half reliability = 0.95); if we removed outlier responses for each participant that deviated >2 standard deviations from the mean response across all eight runs for each target group on each scale (<5% of total responses), results were consistent with those reported below. For mean behavioral ratings of each group on each scale, see Table 1. As predicted, relative to high-status groups, the low-status groups were significantly dehumanized. We also observed that low-status groups were more disliked, seen as significantly more dissimilar from the self, and perceived to be marginally more homogeneous (see Table 2). Participants also rated animal targets significantly lower on the humanization (and heterogeneity) scales than both high- and low-status groups ($ts > 2.5$, $ps < 0.02$), whereas animals were not significantly distinguished from human groups on either the liking or similarity scales ($ts < 1.8$, $ps > 0.09$).[5]

Similar results were obtained in a behavioral study using an independent sample who evaluated the same groups using the same scales (see online supplementary materials).

Neuroimaging Results

Whole brain regression analyses. Our primary analyses were focused on determining whether rating the dehumanization of groups generated distinct patterns of neural activation relative to providing ratings of the same groups on other conceptually related dimensions. To do this, we looked for regions in the whole brain that were significantly correlated—either positively or negatively[6]—with any of the scales (e.g., regions where activity was positively correlated with humanization ratings, and regions where activity was negatively correlated with humanization ratings), resulting in eight parametric analyses.

Aside from discrete visual cortical responses resulting from the experimental design (see Figure S1), dehumanization of target groups was associated with activity in four brain regions: dorsomedial precuneus (PC), two regions in left inferior parietal cortex (IPC), and one region in left inferior frontal cortex (IFC). Liking of target groups was associated with activity in a region in posterior cingulate cortex (PCC). These regions were all subjected to follow up ROI analyses. No region showed a significant response, corrected for multiple comparisons, in the other six parametric analyses: humanization, disliking, similarity, dissimilarity, heterogeneity, or homogeneity. However, a single region in the MPFC was

sensitive to parametric ratings of similarity to self at a more lenient threshold ($p < .001$, k > 10). This region was similar to the one identified by Harris and Fiske (2006) and Krendl (2016) to be associated with passively viewing low-warmth/low-competence individuals (e.g., homeless people, drug addicts) versus higher warmth/competence individuals (e.g., lawyers), and the region identified by Mitchell et al. (2006) to be sensitive to judging similar versus dissimilar others, and was therefore included in the ROI analyses (Figure 3; Table 3).

Region of interest analyses. In a first set of ROI analyses, we wanted to determine if the regions that responded parametrically to a rating scale were *selective* for that scale controlling for potentially related judgments. To examine the specificity of the parametric neural responses in each ROI, we performed a series of multiple regression analyses using the ratings for each group on each scale across all participants. For each of the ROIs, we calculated the mean brain response (in beta) for each of the evaluations (four scales by 10 targets) per individual, and compared these neural responses with the behavioral score for each of the evaluations. For every evaluation scale we built a multilevel model predicting the brain response while engaging in each type of evaluation using all of the behavioral scores, with subject as a random factor in the model. To determine if a brain region that correlated significantly with a rating scale did so not only when controlling for all other scale ratings, but also significantly more *strongly* than all other scales, we then compared the strength of correlations between all scales within each ROI.

The whole brain analysis revealed four brain regions that were parametrically sensitive to dehumanization ratings. In the left IFC, neural responses were selectively sensitive to behavioral ratings on the Ascent Dehumanization scale: The association between behavioral ratings and brain responses remained significant when controlling for behavioral ratings on all other scales ($p < .001$). Moreover, none of the other behavioral ratings were significantly associated with neural responses in this region ($ps > .29$). This pattern of responses was present only when making blatant dehumanization judgments: Ascent dehumanization ratings did not predict neural responses in this region when participants made judgments on any of the other scales (nor did ratings on any other scale, $ps > .07$). Additionally, when making dehumanization judg-

---

[5] Note that similar mean ratings for humans and animals on similarity judgments were driven mostly by people's high ratings of similarity with puppies. Given that dogs are a beloved pet to many, and likely familiar to all Americans, finding that dogs are viewed as more similar to participants than Muslims and Gypsies (despite being rated as lower on the humanization scale) stood to reason.

[6] Given that the human brain contains both excitatory and inhibitory neuronal populations, it seemed equally plausible to us that a brain region could be parametrically sensitive to increasing ratings of humanization, liking, similarity, or heterogeneity, which could activate downstream prosocial actions and/or inhibit antisocial actions, or parametrically sensitive to increasing ratings of *dislike, dehumanization, dissimilarity, or homogeneity*, which could activate downstream anti-social process and/or inhibit prosocial actions. Consistent with this view, previous research has reported activity in brain regions that is greater to similar/humanized/liked/high-status others versus dissimilar/dehumanized/disliked/low-status others (e.g., Mitchell et al., 2006), while other research has shown the reverse pattern in other brain regions (e.g., Cloutier, Ambady, Meagher, & Gabrieli, 2012); Jack et al. (2013) report greater activity to animals/animalistic activity versus humans/humanistic activity in some brain regions, and the reverse in other brain regions, within the same experiments.

Table 1

Average Behavioral Ratings for Each Group on Each Scale

Scale	Group	Mean	SD	Scale	Group	Mean	SD
Ascent (Humanization)	Americans	17.91	2.57	Overlap (Similarity)	Americans	15.46	4.08
	Europeans	17.80	3.26		Europeans	9.74	4.54
	surgeons	18.36	2.84		surgeons	8.06	5.01
	Muslims	16.65	4.74		Muslims	7.38	4.91
	Gypsies	15.51	3.95		Gypsies	5.10	3.28
	homeless	15.80	4.05		homeless	6.63	4.46
	puppies	6.44	4.34		puppies	6.99	4.96
	rats	4.04	3.77		rats	2.88	3.15
	toddlers	11.37	6.88		toddlers	5.19	3.86
	robots	10.95	4.45		robots	8.34	5.63
Feeling Thermometer (Liking)	Americans	15.00	3.48	Heterogeneity	Americans	14.89	5.04
	Europeans	14.05	3.76		Europeans	14.71	4.79
	surgeons	14.37	4.21		surgeons	9.93	4.44
	Muslims	12.08	4.87		Muslims	11.93	5.00
	Gypsies	9.10	3.37		Gypsies	11.37	3.55
	homeless	10.01	4.56		homeless	11.83	3.92
	puppies	15.48	4.22		puppies	11.56	3.58
	rats	5.37	4.13		rats	7.28	4.96
	toddlers	10.21	5.69		toddlers	10.62	5.11
	robots	13.78	5.17		robots	10.90	4.09

ments, the strength of the correlation between neural responses and dehumanization ratings was significantly greater than the correlation between neural responses and ratings of dislike, $t(212) = 12.65$, $p < .001$, and homogeneity, $t(212) = 4.28$, $p = .040$, and marginally stronger than the correlation between neural responses and dissimilarity, $t(212) = 2.97$, $p = .086$.

The relationship between neural responses and ascent dehumanization ratings in the PC remained marginally significant when controlling for ratings on all other scales ($p = .084$); dissimilarity ratings were also associated with neural responses while making dehumanization judgments in this model ($p = .044$), whereas ratings on homogeneity and feeling thermometer were not (ps > .44). Dehumanization ratings correlated (marginally) with neural responses in the PC only when making dehumanization judgments, not with neural responses when making judgments on the other scales (i.e., this association was specific). However, when making dehumanization judgments, the correlation between neural re-

Table 2

Mean Ratings of High-Status Human Groups (Americans, Europeans, Surgeons) and Low-Status Groups (Muslims, Gypsies, Homeless)

Scale/status	Means	95% CI	t-value	p-value	d
Humanization					
High	18.02	[17.04, 19.01]	3.43	.002	.30
Low	15.99	[14.29, 17.69]			
Liking					
High	14.48	[13.21, 15.74]	4.94	<.001	.43
Low	10.39	[9.07, 11.72]			
Overlap					
High	11.09	[9.74, 12.43]	6.46	<.001	.56
Low	6.37	[4.97, 7.77]			
Heterogeneity					
High	13.18	[11.52, 14.83]	1.80	.085	.16
Low	11.71	[10.48, 12.94]			

Note. Ratings were provided on 0–20 scales.

sponses and dehumanization ratings was not significantly stronger in the PC than the correlation between neural responses and ratings of dislike ($p = .11$), dissimilarity ($p = .50$), or homogeneity ($p = .23$).

The two left inferior parietal regions showed specificity to ascent dehumanization ratings similar to the left IFC: The association between behavioral ratings and brain responses remained significant when controlling for behavioral ratings on all other scales (ps < .001), and this pattern of responses was present only when making ascent dehumanization judgments (all other ps > .11). Finally, neural responses were significantly more strongly correlated with ascent dehumanization judgments than with dislike (LIP 1: $t(212) = 4.91$, $p = .028$; LIP 2: $t(212) = 24.99$, $p < .001$) and homogeneity (LIP 1: $t(212) = 6.87$, $p = .009$; LIP 2: $t(212) = 6.36$, $p = .009$), though not dissimilarity (LIP 1: $t(212) = 1.25$, $p = .26$; LIP 2: $t(212) = .05$, $p = .82$).

We also examined the specificity of responses to liking in the PCC. We found that the relationship between neural responses while making feeling thermometer judgments and behavioral liking ratings was significant when controlling for ratings on all other scales ($p = .011$); ratings on the other scales did not predict beta values while making feeling thermometer judgments (ps > .41), and no scale's ratings predicted beta values in this region while making judgments on any of the other scales (ps > .19). When making liking judgments, neural activity was significantly more strongly correlated with liking ratings than with ratings on the Ascent Dehumanization scale ($p = .014$), but showed only a trend toward stronger correlations than homogeneity ratings ($p = .11$) or dissimilarity ratings ($p = .23$).

In a second set of ROI analyses, we sought to determine if brain regions that displayed parametric sensitivity to dehumanization judgments across all groups also showed a pattern of mean activity especially consistent with dehumanization of low-status groups. To do this, we directly examined mean neural responses to the low-status and high-status human groups that are of primary theoretical interest in social psychological research, and compared
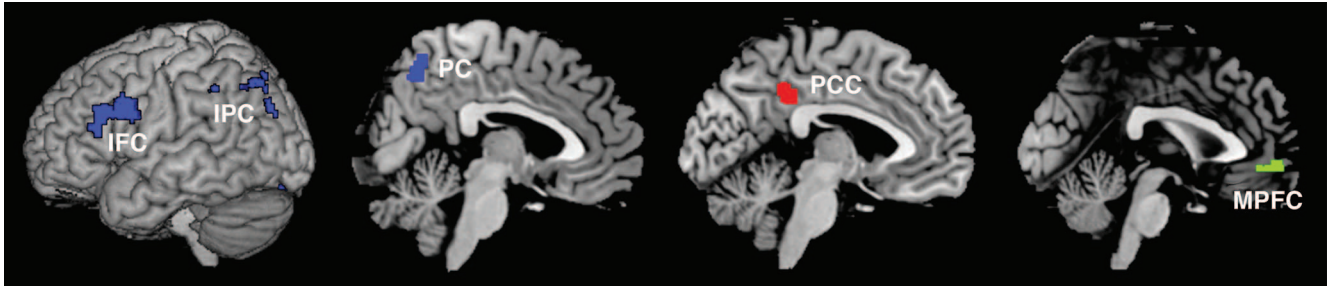
Figure 3. Brain regions where activity is sensitive to parametric ratings of dehumanization (blue), liking (red), and similarity to the self (green). Dehumanization and liking thresholded at $p < .05$, corrected; similarity thresholded at $p < .001$, uncorrected. See the online article for the color version of this figure.

them with mean neural responses to animals (one liked, puppies, and one disliked, rats, but both distinctively nonhuman). We theorized that the strongest evidence for a socially relevant dehumanization-sensitive brain response would be a region that was not only significantly parametrically modulated by dehumanization judgments (controlling for *and* more strongly than potentially overlapping judgments), but also responded more similarly to animals and low-status groups relative to high-status groups (consistent with behavioral research suggesting that low-status groups tend to be more animalistically dehumanized than high-status groups).

To examine this, we performed a 4 scale $\times$ 3 group within-subjects ANOVA in each ROI that proved sensitive to ascent dehumanization. For any regions that showed a Scale $\times$ Group interaction, we performed follow-up univariate ANOVAs and planned paired-samples t tests across groups (high-status human, low-status human, animal) within each scale. For each of these regions, we found a significant Scale $\times$ Group interaction ($Fs > 4.3$, $ps < 0.001$, $\eta^2s > 0.15$), suggesting that the pattern of group responses differed significantly across scales. Follow-up analyses revealed that each of the regions also showed a significant univariate effect across groups only when making dehumanization judgments ($Fs > 8.1$, $ps < 0.001$, $\eta^2s > 0.26$; all other scales for all

Table 3
Regions Parametrically Sensitive to Each of the Rating Scales

Scale/region	x	y	z	size	t
Like					
PCC	−8	−40	34	182	5.96
Similar					
MPFC	0	52	−4	39	4.09[a]
Dehumanize					
PC	4	−62	46	337	4.82
L IFC	−44	6	30	461	7.00
L IPC 1	−38	−44	36	110	6.79
L IPC 2	−30	−76	32	331	5.21

Note. PCC = posterior cingulate cortex; MPFC = medial prefrontal cortex; PC = dorsomedial precuneus; L IPC = left inferior parietal cortex; L IFC = left inferior frontal cortex.
No other regions were sensitive to parametric ratings at these thresholds, and no regions were sensitive to parametric ratings of humanization, dislike, dissimilarity, homogeneity or heterogeneity.
[a] Significant at $p < .001$, uncorrected (all others significant at $p < .05$, corrected).

four ROIs: $Fs < 2.9$, $ps > 0.05$). Finally, planned paired-samples t tests were performed to determine whether the specific pattern of responses to the three groups (high-status, low-status, animals) was consistent with a general human versus animal distinction (i.e., animals > [low-status = high-status]), or if it further distinguished between high- and low-status human groups, with low-status groups showing an activation profile closer to that for animals (i.e., animals ≥ low-status > high-status).

We found that when making dehumanization judgments, both of the left IPC regions responded more strongly to animals (left IPC region 1: $M = 0.231$, $SD = 0.323$; left IPC region 2: $M = 0.806$, $SD = 0.828$) than to human groups that were both low-status (left IPC region 1: $M = 0.074$, $SD = 0.344$; left IPC region 2: $M = 0.617$, $SD = 0.750$; $ts > 3.1$, $ps < 0.015$, $ds > 0.65$) and high-status (left IPC region 1: $M = 0.021$, $SD = 0.297$; left IPC region 2: $M = 0.521$, $SD = 0.704$; $ts > 4.4$, $ps < 0.007$, $ds > 0.90$), but the human groups were indistinguishable from each other ($ts < 2.1$, $ps > 0.150$). Therefore, the left IPC regions distinguished between animals and humans specifically when making dehumanization judgments, but did not distinguish low-status and high-status human groups.

The left IFC showed a similar pattern, but also showed a trend—specifically during dehumanization judgments—that distinguished between low- and high-status human groups: neural responses in this region were higher while making dehumanization ratings of animals ($M = 0.499$, $SD = 0.594$) than when rating both high-status ($M = 0.140$, $SD = 0.603$; $t(23) = 6.38$, $p < .001$, $d = 1.11$) and low-status ($M = 0.295$, $SD = 0.603$; $t(23) = 2.82$, $p = .029$, $d = 0.56$) groups, and responses while rating low-status groups were marginally greater than while rating high-status groups, $t(23) = 2.41$, $p = .074$, $d = 0.49$.

The dorsomedial PC showed a pattern of responses similar to the left IFC: neural responses while making dehumanization ratings of low-status groups ($M = 0.475$, $SD = 0.682$) and animals ($M = 0.536$, $SD = 0.619$) that were significantly higher than while making dehumanization judgments about high-status human groups ($M = 0.252$, $SD = 0.495$; low-status vs. high-status: $t(23) = 3.03$, $p = .018$, $d = 0.72$; animals vs. high-status: $t(23) = 4.32$, $p = .001$, $d = 0.95$). Moreover, the neural responses in this region when making judgments about low-status groups and animals were statistically indistinguishable from each other, $t(23) = 0.75$, $p = .459$. Within these four ROIs, no other paired-samples t tests were significant between any groups (animals, low-status,

high-status) on any of the other scales (dislike, dissimilarity, homogeneity; all $ps > .05$; see Figure 4).

The two ROIs that were parametrically responsive to liking (PCC) and similarity (MPFC) showed less sensitivity to the type of group being judged. Because there was no specific prediction about mean responses to animals versus high- and low-status human groups on the liking and similarity scales, we examined each intergroup comparison (high- vs. low status human groups, high-status vs. animals) using 4 Scale × 2 Group ANOVAs, separately for each region. Across both regions, there was no main effect of scale, group or any Scale × Group interactions (all $Fs <$ 2.50, $ps > 0.065$); results were similar if animals were included as a third group and 4 Scale × 3 Groups ANOVAs were performed.

Therefore, across our full set of analyses the response profile that provided the *most* compelling and consistent evidence of a role in dehumanization was the left IFC: Activation in this region during dehumanization trials was parametrically modulated by dehumanization judgments at corrected thresholds controlling for all other judgments; was *more* sensitive to blatant dehumanization judgments than to judgments on all other scales; tracked behavioral ratings of dehumanization only during trials relating to dehumanization; and had mean activation levels that distinguished between high-status human groups, low-status human groups and animals. In comparison, we found evidence suggesting that acti-

vation in the left IPC regions was uniquely parametrically modulated by dehumanization judgments (vs. other judgments), but little evidence that activation in this region differentiated between high- and low-status humans. Mean activation in the dorsomedial PC, differentiated between high- and low-status humans, but its parametric sensitivity controlling for other scale ratings was marginal and was not stronger than for any other judgments. By contrast, a region in the PCC was parametrically sensitive to liking judgments, and a region in the MPFC was parametrically sensitive (at uncorrected thresholds) to similarity judgments. See Table S1 for mean responses to all groups across all scales in each ROI.

Discussion

In the present study, we examined neural responses associated with blatant dehumanization judgments, and compared them with the neural responses associated with judgments on three conceptually related dimensions: dislike, perceived dissimilarity and within-group homogeneity. The distinction between dehumanization and dislike was particularly interesting to us, given previously articulated concerns in the literature about blatant dehumanization being merely a reflection of general antipathy (Haslam & Loughnan, 2014), and the fairly high correlation ($rs \sim .5–.6$) between blatant dehumanization (assessed with the Ascent Dehumanization
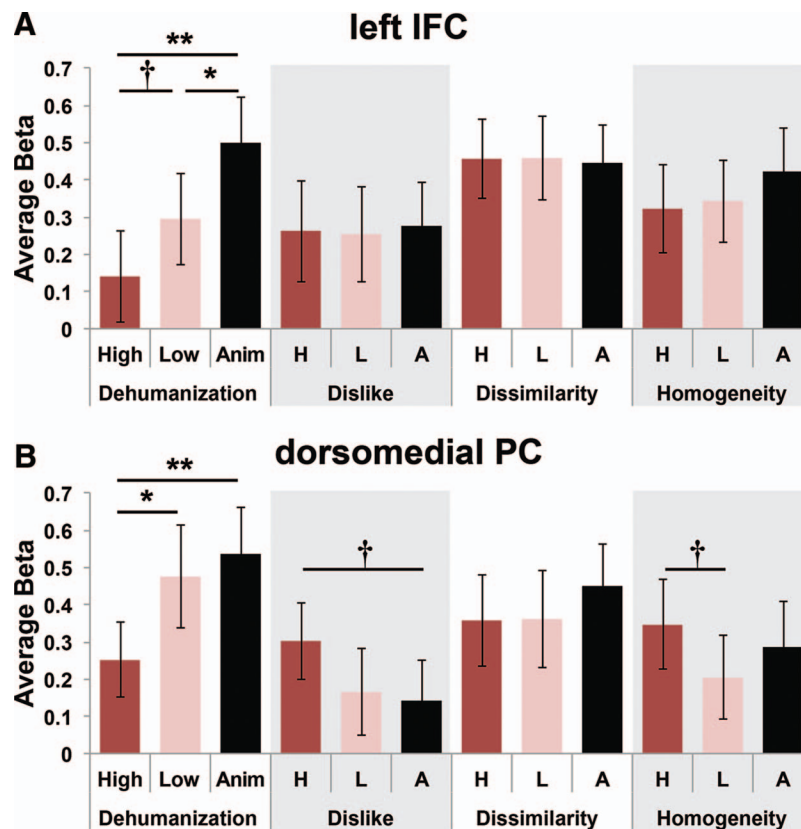

Figure 4. Leave one out region of interest analysis in (A) left inferior frontal cortex and (B) dorsomedial precuneus. In both brain regions, mean responses to the high-status human groups were lower than mean responses to the low-status groups and animals. † $p < .10$. * $p < .05$. ** $p < .005$. See the online article for the color version of this figure.

scale) and dislike (assessed with feeling thermometer) in behavioral research across multiple contexts. A behavioral study we conducted with an independent sample of participants verified that blatant dehumanization was also highly correlated with dissimilarity and homogeneity ratings (see supplementary online study), furthering empirical support for proposed theoretical links between these constructs (e.g., Haslam & Stratemeyer, 2016; Kelman, 1973; McDonald et al., 2017).

Despite the correlations in behavioral ratings, we identified in the neuroimaging study four regions—one in the dorsomedial PC, two in the left IPC and one in the left IFC—where activity was parametrically associated with dehumanization judgments, only while providing dehumanization judgments, and even when controlling for all other scale ratings (marginally significant for dorsomedial PC). Furthermore, activation in left IFC and left IPC was significantly more strongly correlated with dehumanization judgments than with judgments of dislike. The left IFC was particularly notable, in that (unlike the left IPC regions) it further showed greater activity in response to animals and low-status human groups, versus high-status human groups, a response profile consistent with behavioral research suggesting that low-status groups are particularly likely to be likened to animals (e.g., Kteily et al., 2015). Furthermore, and as we discuss further below, the activation we observed in left IFC converges with prior neuroimaging research using a different methodology (i.e., examining spontaneous rather than deliberative dehumanization of targets depicted animalistically; Jack et al., 2013).

On the other hand, a region in the PCC was parametrically associated with liking judgments, even when controlling for all other scale ratings, and only when providing liking judgments. This region was significantly more strongly correlated with liking judgments than with dehumanization judgments. Finally, a single region, in MPFC, was sensitive to similarity judgments, though only at an uncorrected threshold.

These results therefore provide evidence for a double dissociation in brain regions that are parametrically sensitive to evaluations of dehumanization versus dislike (and are suggestive of a triple dissociation between dehumanization, dislike, and dissimilarity). These findings thus provide neural evidence for a longstanding hypothesis that blatant dehumanization is meaningfully different from outgroup affect (Kelman, 1973; Leyens et al., 2000): Certain groups (e.g., children, members of remote tribes like the Pygmies or !Kung) may be liked but still denied cognitive faculties (e.g., capacity for self-control) that are deemed central to full humanity; other groups (e.g., Germans) may be disliked by lower-status Eastern European countries but nevertheless acknowledged to be equally (or even more) "fully evolved" than one's own group (Bruneau et al., 2017); still other groups (e.g., Muslims) may be both disliked *and* dehumanized (Kteily & Bruneau, 2017a; Kteily et al., 2015, 2016). Our work suggests that the behavioral distinction between dehumanization and dislike is reflected in the neural responses to making judgments on each of these dimensions.

Links to Prior Neuroimaging Research on Dehumanization

The evidence reported here identifies brain regions that are specific to deliberative judgments of blatant dehumanization versus judgments on related dimensions. By design, it therefore does

not examine spontaneous dehumanization. Previous research has examined neural responses to passively viewing typically dehumanized groups (i.e., potentially capturing "spontaneous dehumanization" in a naturalistic way), but the design of these studies cannot easily distinguish between dehumanization and related processes. Although it is unlikely that the neural regions active during deliberate and passive dehumanization paradigms would overlap completely, the best evidence for brain regions that may be selectively sensitive to dehumanization in real-world settings would be regions that converge across both the current study and previous research examining spontaneous activation. Here, we consider the degree of convergence between our work and prior research and consider how our findings might help inform that research.

In one study, Harris and Fiske (2006) found decreased MPFC activity in response to images of low-status versus high-status others. The authors interpreted the decreased MPFC activity as a reflection of dehumanization, in so far as decreased MPFC activity might reflect reductions in the tendency to consider the minds of low-status targets. Still, their study was not designed to distinguish between dehumanization and related processes, and the authors acknowledged that their results converged with reports of decreased activity in MPFC when considering dissimilar versus similar others (e.g., Mitchell et al., 2006). Indeed, they suggested that "Future research should parse these two overlapping but nonetheless independent dimensions [dehumanization and dissimilarity]". We provide such evidence here by showing that activity in a region of the MPFC (0 52 −4) very close to the MPFC region identified by Harris and Fiske (2006) (−9 50 −2) tracks with similarity judgments, but not with judgments of blatant dehumanization. That is, our work suggests that the activation observed by Harris and Fiske (2006) does not map on to the deliberative determination of a target as less than fully human, and may instead reflect a more subtle form of dehumanization in which individuals overlook the minds of those they deem to be dissimilar (see also Waytz, Gray, Epley, & Wegner, 2010).

On the other hand, our observation that the left IFC provided the most compelling evidence for a role in deliberative judgments of blatant dehumanization converged with previous neuroimaging studies examining dehumanization using different methodologies. Specifically, Jack et al. (2013) found greater responses in regions of left IFC when people viewed pictures of animals (−46 8 22) or overtly animalistic actions (e.g., drinking from a mud puddle like a dog [−41 −1 31]), versus pictures of humans or human-specific actions (Jack et al., 2013). These peak activations are very close to those reported in the current study (−44 6 30), and neither the current work nor the experiments from Jack et al. (2013) found corresponding activity in right IFC. Thus, both spontaneous dehumanization evoked by passively viewing animalistically dehumanized others (Jack et al., 2013), and deliberative animalistic dehumanization judgments (current data) appear to recruit left-lateralized IFC. This converging evidence across different methods provides a particularly compelling argument for the importance of left IFC to dehumanization, and further suggests its distinctness relative to processes (e.g., in the MPFC) that appear more relevant to judgments of similarity to the self.

What cognitive process might activity in IFC index? That is, what reverse inferences about the psychological mechanism of dehumanization are licensed by our results? In general, reverse inferences from whole-brain group analyses must be treated with

caution, because normalizing and group averaging blur activity from distinct but nearby regions. However, reverse inferences can also be informative, for example by narrowing the possible psychological interpretations of neural responses, and providing hypotheses that can inform future work. To this end, we consider three possible interpretations of IFC activity observed in the current research.

First, the interpretation favored by Jack et al. (2013) is that left IFC represents task-positive network (TPN) activity, which serves to reduce activity in the default mode network (DMN) associated with mentalizing. This would suggest that greater activity in response to animals and low-status humans in the left IFC serves to reduce mentalizing toward these groups, thus dehumanizing them by denying mind (Epley & Waytz, 2010; Waytz & Schroeder, 2014). From this perspective, one would predict simultaneous and opposite parametric sensitivities in DMN/mentalizing regions. In our own work, we did not observe this, although it is possible that this activity could be subthreshold and potentially revealed by studies with greater power.

A second possibility is that greater activity in left IFC in response to low-status versus high-status groups while rating groups on the Ascent Dehumanization scale is due to an increase in cognitive control or self-sanctioning when making provocative and politically incorrect judgments about low-status others. However, this interpretation seems unlikely for two reasons: First, the *highest* responses in these brain regions were observed when making dehumanization judgments about animals, which are unlikely to require cognitive control, or be encumbered by self-presentation concerns. Second, the prior neuroimaging work cited above (Jack et al., 2013) observed left IFC activity in response to passive viewing (rather than deliberative judgments), where self-presentational concerns ought to be much less relevant.

A third possibility is that the selective sensitivity of left IFC to dehumanization judgments (vs. other judgments), and the increased activity in left IFC when making dehumanization judgments about low-status human groups and animals, could represent sensitivities to considerations of social hierarchy. A left-lateralized region of the IFC (-42 17 21) was found to be more active when explicitly judging the relative social status of famous humans, versus judgments about the same targets on other dimensions such as age, gender, or fame (Farrow et al., 2011). And other work found that a partially overlapping region of IFC, also left-lateralized (-42 20 34), responded more strongly when passively viewing images of financially low-status versus high-status individuals (Cloutier et al., 2012). These studies provide converging evidence for the potential relevance of left IFC in processing social status.

Although our interpretation that dehumanization-specific left IFC activity in our study relates to judgments of social standing remains tentative, it is consistent with prior behavioral evidence suggesting the importance of status considerations to blatant dehumanization. Indeed, individuals particularly concerned with maintaining hierarchical differentiation between social groups (i.e., high on social dominance orientation; Ho et al., 2015) are especially likely to rate low-status groups as animals (Kteily et al., 2015).

How might this explanation reconcile with the lack of left IFC activity observed when making judgments about dislike, dissimilarity, and homogeneity? Although status is unlikely to be entirely irrelevant to these judgments, they may depend less specifically on perceptions of social standing, and more on perceptions of targets' warmth, closeness, or familiarity. Even high-status targets might sometimes be rated as dissimilar due primarily to their differing cultural practices, or as homogenous due to our lack of (individuating) contact with them. And high-status is not necessarily determinative of liking either: High-status targets are often stereotyped as cold (Fiske, Cuddy, & Glick, 2007; Judd, James-Hawkins, Yzerbyt, & Kashima, 2005). In fact, recent work shows that lower-status Eastern European countries (Hungary, Czech Republic) readily report dislike of higher-status European countries (e.g., Germany) even while rating them to be just as human (and, at times, even *more* human) than their own groups (Bruneau et al., 2017).

Implications, Limitations, and Future Directions

Beyond the advances it provides, our work also raises additional questions for future research. For one, whereas the design of the current work and that employed by Jack and colleagues (2013) differed both in structure and target, it would be interesting to examine the degree of convergence between neural activation when individuals are consciously judging versus passively viewing the same targets.

It would also be interesting to expand consideration of the *types* of conscious humanity-related judgments individuals make. We focused here on especially blatant judgments—directly likening targets to lower animals—but other work has examined more subtle attributions. For example, research on infrahumanization has investigated the tendency to attribute fewer complex secondary emotions (thought to be uniquely human) to certain groups (i.e., infrahumanization; Leyens et al., 2000), a pattern that can hold in more "everyday" contexts, and even if individuals do not consciously think of the targets as more animalistic. Might it be the case that activation underlying this more subtly dehumanizing judgment converges even more strongly with activation predicting the spontaneous dehumanization of passively viewed targets that are not overtly depicted as animals? Exploring this question might help further delineate the distinction between subtle and blatant dehumanization (see also Kteily & Bruneau, 2017b).

Future work could also examine the activation underlying forms of dehumanization beyond the animalistic dehumanization we focused on here. Haslam (2006) highlighted that we also sometimes *mechanistically* dehumanize targets (e.g., Germans), seeing them not as animals but as rigid automata lacking in "human nature" traits such as curiosity and warmth. Other research highlights that we sometimes *objectify* others (e.g., employees), deindividuating them and treating them as mere tools useful to achieving our instrumental goals (e.g., Orehek & Weaverling, 2017). It would be interesting to consider how activation associated with animalistic dehumanization judgments might be similar to or different from that associated with these other forms of dehumanization. For example, whereas we observed a double-dissociation between activation in regions associated with dislike and animalistic dehumanization judgments, it could be that activation in regions associated with liking judgments is more relevant to whether we see targets as cold robots or warm humans. Similarly, judging targets as a homogenous mass rather than unique individuals may be particularly central to objectification. Examining how

liking, dissimilarity and homogeneity relate to different forms of dehumanization—at both the behavioral and neural levels—could therefore be an interesting line of future research.

Notwithstanding the new questions posed by our work, our findings have important implications. For one, despite any uncertainties about its roots, the activation we observe in the IFC in response to actively dehumanizing low-status groups and animals could have downstream effects on social–cognitive processing (as suggested by Jack et al., 2013), potentially inhibiting concerns about the consequences of harmful actions to these targets. By using—and oftentimes abusing—animals for food, sport, and clothing, we frequently place them outside the scope of our moral concern and show relatively little regard for their suffering. Importantly, this same type of moral exclusion provides the bedrock for theoretical explanations of how humans have committed atrocities against other humans throughout history (Bandura, Barbaranelli, Caprara, & Pastorelli, 1996). The findings reported here suggest one potential neural mechanism that may underlie a shared moral disregard toward both animals and low-status human groups, distinguishing them both from "fully evolved" humans. Future research could examine the connectivity between the brain regions identified here, and activation of brain regions associated with social cognition and moral concern, and the association of activity in these regions and prosocial and antisocial intergroup attitudes and behaviors.

The neural distinction we identify between deliberative dehumanization and dislike also has implications for the potential efficacy of interventions. Efforts to decrease intergroup hostility have almost exclusively focused on affective processes, such as decreasing outgroup coldness or improving outgroup empathy. If prejudice and dehumanization are cognitively distinct processes, it is possible that making a group seem more likable may not override, for example, a tendency to consciously perceive them as primitive or brutish. Interventions aiming to decrease intergroup discrimination rooted in deliberative dehumanization may thus need to more directly target the association between certain groups and animals, or focus on promoting a sense of the target outgroup as civilized, sophisticated, and capable. Future work may help determine if interventions specific to dehumanization versus dislike influence distinct neural responses and predict different intergroup outcomes.

Finally, it is important to acknowledge the limitations of this research. First, although our approach of assessing deliberative judgments of blatant dehumanization has important advantages, it also has certain downsides. Blatant judgments of dehumanization may be particularly subject to self-presentation concerns, with some individuals likely feeling a need to censor their true judgments of certain targets. Although behavioral studies assessing blatant dehumanization using deliberative judgments find that it strongly predicts hostile intergroup behavior (suggesting that it does capture meaningful variance beyond social desirability; see Kteily & Bruneau, 2017b, for a review), avoiding self-presentation concerns entirely is a potential advantage of neuroimaging approaches that examine neural activity during passive viewing; for this reason we find the converging evidence for left IFC activation across paradigms encouraging. Second, as with most neuroimaging studies the statistical inferences would have been bolstered with a larger sample size, and these results are limited to a single participant sample. It could also be useful to examine a larger

sample of groups (see Harris & Fiske, 2011). Lastly, and perhaps most notably, although we had a clear a priori prediction that judgments of dehumanization would generate unique neural activity relative to judgments of dislike, our predictions about which brain regions would be activated by dehumanization or dislike were more limited. Future work could use the neural activity identified here as regions of interest for confirmatory analysis.

Conclusion

Recent empirical work has demonstrated that blatant dehumanization is a relevant psychological process associated with contemporary intergroup hostility. Here, we identify brain regions that are parametrically sensitive to judgments of blatant dehumanization, but insensitive to the conceptually related judgments of dislike, dissimilarity, and perceived homogeneity. We also identify a region that is sensitive to dislike, but insensitive to dehumanization. Together, this provides neural evidence for a cognitive dissociation between blatant dehumanization and dislike. Our work provides especially compelling evidence for the distinct role of left IFC activation in dehumanization, converging with evidence from other neuroimaging research on dehumanization using different methodologies. Taken together, our work speaks to a long-standing concern in the dehumanization literature that blatant dehumanization and blatant dislike are conflated, informs previous research on spontaneous dehumanization, and identifies a set of brain regions that can be examined in future work aimed at dissecting the cognitive processes involved in blatantly dehumanizing others.

References

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. http://dx .doi.org/10.1037/0022-3514.71.2.364

Bandura, A., Underwood, B., & Fromson, M. E. (1975). Disinhibition of aggression through diffusion of responsibility and dehumanization of victims. *Journal of Research in Personality, 9,* 253–269. http://dx.doi .org/10.1016/0092-6566(75)90001-X

Bruneau, E., & Kteily, N. (2017). The enemy as animal: Symmetric dehumanization during asymmetric warfare. *PLoS ONE, 12,* e0181422. http://dx.doi.org/10.1371/journal.pone.0181422

Bruneau, E., Kteily, N., & Laustsen, L. (2017). The role of Blatant dehumanization in rejection of and hostility towards Muslim refugees during the European "Refugee Crisis." *European Journal of Social Psychology.* Advance online publication. http://dx.doi.org/10.1002/ejsp .2357

Büchel, C., Holmes, A. P., Rees, G., & Friston, K. J. (1998). Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. *NeuroImage, 8,* 140–148. http://dx.doi.org/10.1006/ nimg.1998.0351

Cloutier, J., Ambady, N., Meagher, T., & Gabrieli, J. D. E. (2012). The neural substrates of person perception: Spontaneous use of financial and moral status knowledge. *Neuropsychologia, 50,* 2371–2376. http://dx .doi.org/10.1016/j.neuropsychologia.2012.06.010

Cuddy, A., Fiske, S., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology, 40,* 61–149.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science, 293,* 2470–2473. http://dx.doi.org/10.1126/science.1063414

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America, 113,* 7900–7905. http://dx.doi.org/10.1073/pnas.1602413113

Epley, N., & Waytz, A. (2009). Mind Perception. In S. T. Fiske, D. T. Gilbert, & G. Lindsay (Eds.), *The Handbook of Social Psychology* (5th ed., pp. 498–541). New York, NY: Wiley.

Esterman, M., Tamber-Rosenau, B. J., Chiu, Y.-C., & Yantis, S. (2010). Avoiding non-independence in fMRI data analysis: Leave one subject out. *NeuroImage, 50,* 572–576. http://dx.doi.org/10.1016/j.neuroimage.2009.10.092

Farrow, T. F., Jones, S. C., Kaylor-Hughes, C. J., Wilkinson, I. D., Woodruff, P. W., Hunter, M. D., & Spence, S. A. (2011). Higher or lower? The functional anatomy of perceived allocentric social hierarchies. *NeuroImage, 57,* 1552–1560. http://dx.doi.org/10.1016/j.neuroimage.2011.05.069

Fiske, A. P., & Rai, T. S. (2014). *Virtuous violence: Hurting and killing to create, sustain, end, and honor social relationships.* New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781316104668

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11,* 77–83. http://dx.doi.org/10.1016/j.tics.2006.11.005

Fiske, S., Cuddy, A., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82,* 878–902.

Goff, P. A., Eberhardt, J. L., Williams, M. J., & Jackson, M. C. (2008). Not yet human: Implicit knowledge, historical dehumanization, and contemporary consequences. *Journal of Personality and Social Psychology, 94,* 292–306. http://dx.doi.org/10.1037/0022-3514.94.2.292

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *science, 315,* 619–620.

Haddock, G., Zanna, M. P., & Esses, V. M. (1993). Assessing the structure of prejudicial attitudes: The case of attitudes toward homosexuals. *Journal of Personality and Social Psychology, 65,* 1105–1118. http://dx.doi.org/10.1037/0022-3514.65.6.1105

Harel, A., Kravitz, D. J., & Baker, C. I. (2014). Task context impacts visual object processing differentially across the cortex. *Proceedings of the National Academy of Sciences of the United States of America, 111,* E962–E971. http://dx.doi.org/10.1073/pnas.1312567111

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme out-groups. *Psychological Science, 17,* 847–853. http://dx.doi.org/10.1111/j.1467-9280.2006.01793.x

Harris, L. T., & Fiske, S. T. (2007). Social groups that elicit disgust are differentially processed in mPFC. *Social Cognitive and Affective Neuroscience, 2,* 45–51. http://dx.doi.org/10.1093/scan/nsl037

Harris, L. T., & Fiske, S. T. (2011). Dehumanized perception: A psychological means to facilitate atrocities, torture, and genocide? *The Journal of Psychology, 219,* 175–181. http://dx.doi.org/10.1027/2151-2604/a000065

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review, 10,* 252–264. http://dx.doi.org/10.1207/s15327957pspr1003_4

Haslam, N., & Loughnan, S. (2014). Dehumanization and infrahumanization. *Annual Review of Psychology, 65,* 399–423. http://dx.doi.org/10.1146/annurev-psych-010213-115045

Haslam, N., & Stratemeyer, M. (2016). Recent research on dehumanization. *Current Opinion in Psychology, 11,* 25–29. http://dx.doi.org/10.1016/j.copsyc.2016.03.009

Heinzel, A., Bermpohl, F., Niese, R., Pfennig, A., Pascual-Leone, A., Schlaug, G., & Northoff, G. (2005). How do we modulate our emotions? Parametric fMRI reveals cortical midline structures as regions specifically involved in the processing of emotional valences. *Cognitive Brain Research, 25,* 348–358. http://dx.doi.org/10.1016/j.cogbrainres.2005.06.009

Ho, A. K., Sidanius, J., Kteily, N., Sheehy-Skeffington, J., Pratto, F., Henkel, K. E., . . . Stewart, A. L. (2015). The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new SDO$_7$ scale. *Journal of Personality and Social Psychology, 109,* 1003–1028. http://dx.doi.org/10.1037/pspi0000033

Hodson, G., Kteily, N., & Hoffarth, M. (2014). Of filthy pigs and subhuman mongrels: Dehumanization, disgust, and intergroup prejudice. *Testing, Psychometrics, Methodology in Applied Psychology, 21,* 267–284.

Jack, A. I., Dawson, A. J., & Norr, M. E. (2013). Seeing human: Distinct and overlapping neural signatures associated with two forms of dehumanization. *NeuroImage, 79,* 313–328. http://dx.doi.org/10.1016/j.neuroimage.2013.04.109

Jahoda, G. (2014). An anthropological history of dehumanization from late-18th to mid-20th centuries. In P. G. Bain, J. Vaes, & J. P. Leyens (Eds.), *Humanness and dehumanization* (pp. 13–33). New York, NY: Routledge.

Jardina, A., & Piston, S. (2016, February). *Dehumanization of black people motivates white support for punitive criminal justice policies.* In Annual Meeting of the American Political Science Association in Philadelphia, PA.

Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology, 89,* 899–913. http://dx.doi.org/10.1037/0022-3514.89.6.899

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience, 17,* 4302–4311.

Kelman, H. C. (1973). Violence without moral restraint: Reflections on the dehumanization of victims and victimizers. *Journal of Social Issues, 29,* 25–61. http://dx.doi.org/10.1111/j.1540-4560.1973.tb00102.x

Kikyo, H., Ohki, K., & Miyashita, Y. (2002). Neural correlates for feeling-of-knowing: An fMRI parametric analysis. *Neuron, 36,* 177–186. http://dx.doi.org/10.1016/S0896-6273(02)00939-X

Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *The Journal of Neuroscience, 21,* RC159.

Krendl, A. C. (2016). An fMRI investigation of the effects of culture on evaluations of stigmatized individuals. *NeuroImage, 124,* 336–349. http://dx.doi.org/10.1016/j.neuroimage.2015.08.030

Kriegeskorte, N., Lindquist, M. A., Nichols, T. E., Poldrack, R. A., & Vul, E. (2010). Everything you never wanted to know about circular analysis, but were afraid to ask. *Journal of Cerebral Blood Flow and Metabolism, 30,* 1551–1557. http://dx.doi.org/10.1038/jcbfm.2010.86

Kteily, N., & Bruneau, E. (2017a). Backlash: The politics and real-world consequences of minority group dehumanization. *Personality and Social Psychology Bulletin, 43,* 87–104. http://dx.doi.org/10.1177/0146167216675334

Kteily, N. S., & Bruneau, E. (2017b). Darker demons of our nature: The need to (re) focus attention on blatant forms of dehumanization. *Current Directions in Psychological Science, 26,* 487–494. http://dx.doi.org/10.1177/0963721417708230

Kteily, N., Bruneau, E., Waytz, A., & Cotterill, S. (2015). The ascent of man: Theoretical and empirical evidence for blatant dehumanization. *Journal of Personality and Social Psychology, 109,* 901–931. http://dx.doi.org/10.1037/pspp0000048

Kteily, N., Hodson, G., & Bruneau, E. (2016). They see us as less than human: Metadehumanization predicts intergroup conflict via reciprocal dehumanization. *Journal of Personality and Social Psychology, 110,* 343–370. http://dx.doi.org/10.1037/pspa0000044

Lammers, J., & Stapel, D. A. (2011). Power increases dehumanization. *Group Processes & Intergroup Relations, 14,* 113–126. http://dx.doi.org/10.1177/1368430210370042

Leyens, J., Paladino, P. M., Rodriguez-Torres, R., Vaes, J., Demoulin, S., Rodriguez-Perez, A., & Gaunt, R. (2000). The emotional side of prejudice: The attribution of secondary emotions to ingroups and outgroups. *Personality and Social Psychology Review, 4,* 186–197. http://dx.doi.org/10.1207/S15327957PSPR0402_06

Lindén, M., Björklund, F., & Bäckström, M. (2016). What makes authoritarian and socially dominant people more positive to using torture in the war on terrorism? *Personality and Individual Differences, 91,* 98–101. http://dx.doi.org/10.1016/j.paid.2015.11.058

Malgady, R. G. (1987). Contrasting part correlations in regression models. *Educational and Psychological Measurement, 47,* 961–965. http://dx.doi.org/10.1177/0013164487474011

Martínez, R., Rodríguez-Bailón, R., & Moya, M. (2012). Are they animals or machines? Measuring dehumanization. *The Spanish Journal of Psychology, 15,* 1110–1122. http://dx.doi.org/10.5209/rev_SJOP.2012.v15.n3.39401

McDonald, M., Porat, R., Yarkoney, A., Reifen Tagar, M., Kimel, S., Saguy, T., & Halperin, E. (2017). Intergroup emotional similarity reduces dehumanization and promotes conciliatory attitudes in prolonged conflict. *Group Processes & Intergroup Relations, 20,* 125–136. http://dx.doi.org/10.1177/1368430215595107

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron, 50,* 655–663. http://dx.doi.org/10.1016/j.neuron.2006.03.040

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping, 15,* 1–25. http://dx.doi.org/10.1002/hbm.1058

Orehek, E., & Weaverling, C. G. (2017). On the nature of objectification: Implications of considering people as means to goals. *Perspectives on Psychological Science, 12,* 719–730. http://dx.doi.org/10.1177/1745691617691138

Pacilli, M. G., Roccato, M., Pagliaro, S., & Russo, S. (2016). From political opponents to enemies? The role of perceived moral distance in the animalistic dehumanization of the political outgroup. *Group Processes & Intergroup Relations, 19,* 360–373. http://dx.doi.org/10.1177/1368430215590490

Park, B., & Judd, C. M. (1990). Measures and models of perceived group variability. *Journal of Personality and Social Psychology, 59,* 173–191. http://dx.doi.org/10.1037/0022-3514.59.2.173

Rainville, P., Duncan, G. H., Price, D. D., Carrier, B., & Bushnell, M. C. (1997). Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science, 277,* 968–971. http://dx.doi.org/10.1126/science.277.5328.968

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology, 16,* 235–239. http://dx.doi.org/10.1016/j.conb.2006.03.001

Saxe, R., & Powell, L. J. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science, 17,* 692–699. http://dx.doi.org/10.1111/j.1467-9280.2006.01768.x

Schubert, T. W., & Otten, S. (2002). Overlap of self, ingroup, and outgroup: Pictorial measures of self-categorization. *Self and Identity, 1,* 353–376. http://dx.doi.org/10.1080/152988602760328012

Smith, D. L. (2011). *Less than human: Why we demean, enslave, and exterminate others.* New York, NY: Macmillan.

Spunt, R. P., Falk, E. B., & Lieberman, M. D. (2010). Dissociable neural systems support retrieval of how and why action knowledge. *Psychological Science, 21,* 1593–1598. http://dx.doi.org/10.1177/0956797610386618

Spunt, R. P., Satpute, A. B., & Lieberman, M. D. (2011). Identifying the what, why, and how of an observed action: An fMRI study of mentalizing and mechanizing during action observation. *Journal of Cognitive Neuroscience, 23,* 63–74. http://dx.doi.org/10.1162/jocn.2010.21446

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences, 14,* 383–388. http://dx.doi.org/10.1016/j.tics.2010.05.006

Waytz, A., & Schroeder, J. (2014). Overlooking other: Dehumanization by commission and omission. *Testing, Psychometrics, Methodology in Applied Psychology, 21,* 251–266.

Willems, R. M., Hagoort, P., & Casasanto, D. (2010). Body-specific representations of action verbs: Neural evidence from right- and left-handers. *Psychological Science, 21,* 67–74. http://dx.doi.org/10.1177/0956797609354072

Yarkoni, T., Barch, D. M., Gray, J. R., Conturo, T. E., & Braver, T. S. (2009). BOLD correlates of trial-by-trial reaction time variability in gray and white matter: A multi-study fMRI analysis. *PLoS ONE, 4,* e4257. http://dx.doi.org/10.1371/journal.pone.0004257