# Political (Meta-)Dehumanization in Mental Representations: Divergent Emphases in the Minds of Liberals Versus Conservatives

## Christopher D. Petsko[1] (iD) and Nour S. Kteily[2]

## Abstract

We conducted two reverse-correlation studies, as well as two pilot studies reported in the online supplement (total $N$ = 1,411), on the topics of (a) whether liberals and conservatives differ in the types of dehumanization that they cognitively emphasize when mentally representing one another, and if so, (b) whether liberals and conservatives are sensitive to how they are represented in the minds of political outgroup members. Results suggest that partisans indeed differ in the types of dehumanization that they cognitively emphasize when mentally representing one another: whereas conservatives' dehumanization of liberals emphasizes immaturity (vs. savagery), liberals' dehumanization of conservatives more strongly emphasizes savagery (vs. immaturity). In addition, results suggest that partisans may be sensitive to how they are represented. That is, partisans' *meta-representations*—their representations of how the outgroup represents the ingroup—appear to accurately index the relative emphases of these two dimensions in the minds of political outgroup members.

Political polarization is a pressing social problem in the United States. Polling data from the American National Election Studies suggest that the degree to which liberals and conservatives dislike each other is at an all-time high (Iyengar et al., 2019). Moreover, animosity between political groups is growing faster in the United States than it is in other Western democratic nations (Boxell et al., 2023). In fact, liberals and conservatives have become so divided in the United States that social scientists have started to describe this division as sectarian—that is, as so extreme that it is psychologically akin to a religious division (Finkel et al., 2020). Indeed, research suggests that liberals and conservatives have such strong faith in the moral superiority of their own political groups that they have come to view political outgroup members as alien (Ahler & Sood, 2018), as untrustworthy (Druckman et al., 2022), and in many cases, as sub-human (Cassese, 2021; Martherus et al., 2021; Pacilli et al., 2016).

The tendency to regard political outgroup members as sub-human—referred to throughout this article as *political dehumanization*—is troubling not only because of its direct associations with intergroup hostility (e.g., Kteily et al., 2015) but also because of the effects that it has on those who perceive that they are being dehumanized. Recent research suggests that in general, both liberals and conservatives

expect to be blatantly dehumanized by political outgroup members (Landry et al., 2021; Moore-Berg et al., 2020) and that expecting to be blatantly dehumanized in turn predicts a range of negative outcomes. For example, greater expectation of being blatantly dehumanized by political outgroup members—referred to throughout this article as *political meta-dehumanization*—is correlated not just with a reciprocal tendency to blatantly dehumanize the outgroup in return (Kteily et al., 2016; Landry et al., 2021), but likewise with a tendency to support behaviors that subvert the democratic system (e.g., redrawing voting districts to advantage one's own political party, even if doing so is technically illegal: Moore-Berg et al., 2020). Thus, political dehumanization and political meta-dehumanization are consequential psychological forces—so much so that a failure to circumvent these forces could undermine U.S. democracy itself.

[1]The University of North Carolina at Chapel Hill, USA
[2]Northwestern University, Evanston, IL, USA

**Corresponding Author:**
Christopher D. Petsko, Kenan-Flagler Business School, The University of North Carolina at Chapel Hill, 300 Kenan Center Drive, Chapel Hill, NC 27599, USA.
Email: christopher_petsko@kenan-flagler.unc.edu

Of course, being able to circumvent political (meta-) dehumanization requires understanding political (meta-) dehumanization. The focus of the present article, therefore, is on adding to scientific understanding of these phenomena, focusing on two general research questions. First, do liberals and conservatives differ in the types of dehumanization that they cognitively emphasize when mentally representing one another? Second, are liberals and conservatives sensitive to *how* they are dehumanized by political outgroup members?

The reason for focusing on the former question—of whether liberals and conservatives differ in the types of dehumanization that they cognitively emphasize—is that distinct types of dehumanization are implicated in distinct behavioral responses (e.g., different ways of expressing intergroup animosity; Andrighetto et al., 2014; Haslam & Loughnan, 2014; Kteily & Landry, 2022). For example, one distinction in dehumanization research contrasts "animalistic" with "mechanistic" dehumanization. Whereas animalistic dehumanization is often associated with aggressive responses, mechanistic dehumanization is more likely to be associated with treating a target as inert or as a means to an end (Haslam & Loughnan, 2014). Relatedly, distinct types of dehumanization may necessitate different types of psychological interventions—a target likened to an animal might benefit more from an intervention that emphasizes their intellect whereas a target likened to a machine might benefit more from one that emphasizes their capacity for emotion. This same logic extends beyond the distinction between animalistic and mechanistic dehumanization to any context in which targets might be dehumanized in distinct ways. If researchers wish to identify the downstream consequences of political dehumanization, and to perhaps curb it altogether, they stand to benefit from understanding if and how dehumanization differs between liberals and conservatives.

The reason for focusing on the latter question—of whether liberals and conservatives are sensitive to how they are dehumanized—is to gain purchase on information that might be useful for correcting inaccurate meta-perceptions. Generally, the research literature on political meta-dehumanization suggests that liberals and conservatives overestimate how dehumanized they are in the minds of political outgroup members (Landry et al., 2021; Moore-Berg et al., 2020). However, it may be the case that liberals and conservatives overestimate certain types of dehumanization, but not others. Knowing which elements of partisans' meta-perceptions are most inaccurate can help researchers to tailor interventions aimed at increasing meta-perceptual accuracy—which has itself been implicated in reducing animosity between political groups (Lees & Cikara, 2020).

## Types of Dehumanization in the Minds of Liberals Versus Conservatives

Social scientists have tended to measure political dehumanization as a unitary construct. Although the measure of choice varies from study to study—at times capturing

animalistic dehumanization (e.g., Pacilli et al., 2016), for example, and at other times capturing perceived "evolvedness" of partisans (e.g., Landry et al., 2021; Moore-Berg et al., 2020)—participants are typically asked to reflect on their global impression that partisans are human or inhuman, but are not required to specify which particular attributes make those partisans human or inhuman. The use of unitary measures in the context of political dehumanization may limit our understanding of the phenomenon. For one, there are in principle several distinct *reasons* why two groups may be thought of all less than fully human, a fact obscured by unitary measures. Moreover, in the specific context of partisan perceptions, there is good reason to believe that liberals and conservatives may dehumanize each other based on distinct sets of attributes. For example, research on subtle dehumanization (see Kteily & Landry, 2022, for a review) has revealed that liberals tend to be denied human-uniqueness characteristics related to maturity (e.g., *organized*, *thorough*, *polite*), whereas conservatives tend to be denied human-nature characteristics related to warmth (e.g., *fun-loving*, *sociable*, and *trusting*; Crawford et al., 2013).

The present article expands on this possibility by suggesting that liberals and conservatives may cognitively emphasize the dimensions of *immaturity* and *savagery* to differing degrees while dehumanizing one another. This article's emphasis on immaturity-based and savagery-based dehumanization, respectively, was chosen for a few reasons. A first and key reason is empirical; the results of a pilot study (see Study S1 of the online supplement) revealed that immaturity- and savagery-based dehumanization are (a) empirically distinct from one another and (b) emphasized to differing degrees in the minds of liberals versus conservatives. In particular, this study revealed that the relative emphasis of savagery (vs. immaturity) was more pronounced among liberals' representations of conservatives than among conservatives' representations of liberals. Second, these dimensions were chosen because they bear conceptual similarity to the aforementioned dimensions of human-nature and human-uniqueness capacities, respectively, which have been shown to be emphasized to differing degrees in the minds of liberals versus conservatives (Crawford et al., 2013). Third, the dimensions of immaturity and savagery are conceptually aligned with cultural stereotypes about partisans in the United States. Consider the U.S. cultural stereotype that "conservatives think liberals are stupid, and liberals think conservatives are evil" (Krauthammer, 2002; as cited in Hartman et al., 2023). Such sentiments accord with a cultural proclivity conservatives may have to dehumanize liberals as seeming too immature for their own good and that liberals may have to dehumanize conservatives as seeming cold-hearted and savage.[1]

The present research examines whether these two types of dehumanization, termed *immaturity-based* and *savagery-based* dehumanization, are indeed emphasized to different degrees in the minds of liberals versus conservatives when

thinking about political outgroup members. In particular, the present research examines whether liberals tend to dehumanize conservatives more along the dimension of savagery than immaturity, and whether conservatives tend to do the reverse—that is, dehumanize liberals more along the dimension of immaturity than savagery. Finally, the present research diverges from prior research in that it examines political dehumanization unobtrusively (i.e., implicitly). In particular, the present work relies on a technique known as reverse-correlation image classification (Dotsch & Todorov, 2012), which enables researchers to estimate how it is that groups of people mentally represent one another.

### Reverse Correlation as a Means of Capturing Political (Meta-)Dehumanization

Recent research suggests that reverse correlation can be a viable tool for measuring dehumanization in an unobtrusive, relatively implicit manner (e.g., Kunst et al., 2018). Reverse correlation is a technique in which researchers ask participants to view hundreds of pairs of faces, and in which researchers ask participants to select the face in each pair that most closely approximates their representation of a particular target group (e.g., their representation of what a *liberal* looks like). After the task is complete, researchers can then take the faces that participants chose and average them into a morphed composite image—an image that indexes how it is that participants mentally represent the target group, on average. Recent research has shown that mental representations, as indexed by reverse correlation, are indeed sensitive to detecting dehumanization. For example, when lay perceivers are shown U.S. adults' mental representations of Arabs, perceivers readily recognize these representations as dehumanizing (e.g., as looking "savage" and "un-evolved"; Petsko et al., 2021).

In addition, this past work suggests that mental representations (as indexed by reverse correlation) can pick up on dehumanization into which participants have little or no awareness (Petsko et al., 2021). The benefit of relying on reverse correlation, then, is that it can measure what is in the minds of liberals and conservatives in ways that might not be capturable from self-report measures alone. This feature of reverse correlation is desirable, as the great majority of what scientists know about political dehumanization has come from studies that rely exclusively on self-report measures (e.g., Cassese, 2021; Martherus et al., 2021; Moore-Berg et al., 2020). In the present work, we examine whether reverse correlation can be used to (a) detect different types of dehumanization that might exist in the minds of liberals versus conservatives, respectively, and if so, whether it can be used to (b) weigh in on the question of whether liberals and conservatives are sensitive to how they are dehumanized by political outgroup members. Notably, the use of reverse correlation to assess sensitivity to how one is dehumanized—to assess what we later refer to as *meta-representations* in

particular—stands on its own as a novel contribution of the present analysis.

## The Present Studies

We conducted two reverse-correlation studies ($N = 902$), as well as two pilot studies reported in the online supplement ($N = 509$), on the topics of (a) whether liberals and conservatives differ in the types of dehumanization that they cognitively emphasize when mentally representing one another, and if so, (b) whether liberals and conservatives are sensitive to how they are represented in the minds of political outgroup members. We report all measures and exclusions, but we note that Study 2 contained an additional condition ($n = 195$) that is reported in the online supplement rather than in the main text.[2] In addition, we note that although predictions and analysis plans were pre-registered in advance of data collection, the analyses reported in this article should be regarded as exploratory, as they differ from what was pre-registered.[3] This article reports raw effect sizes as mean differences ($M_{diffs}$) that are encompassed by 95% confidence intervals (CIs), and it reports standardized effect sizes as standard betas ($\beta$s). Statistical power for detecting main effects and interactions will be presented in the Results sections of each study. Data files, codebooks, survey materials, analysis scripts, and pre-registration documentation associated with this article are available on the Open Science Framework (OSF) website: https://osf.io/483v7/.

## Study 1

Study 1 was designed to examine whether liberals and conservatives differ in the types of dehumanization that they cognitively emphasize when mentally representing one another. In particular, Study 1 examined whether liberals' mental representations of conservatives more strongly emphasize *savagery-based* dehumanization than *immaturity-based* dehumanization, and whether conservatives' mental representations of liberals more strongly emphasize the reverse. As noted previously, the decision to home in on these two types of dehumanization—rather than to measure political dehumanization as a more unitary construct (e.g., Moore-Berg et al., 2020; Pacilli et al., 2016)—was motivated by the results of a pilot study that revealed (a) that conservatives and liberals tend to dehumanize each other along two distinct dimensions, and (b) that liberals and conservatives tend to cognitively emphasize these dimensions to different degrees (see Study S1 of the online supplement). However, and as noted previously, the idea that partisans might exhibit distinct types of dehumanization also bears similarity to research on subtle dehumanization, which has revealed that liberals (vs. conservatives) tend to be denied human-uniqueness characteristics related to maturity (e.g., *organized*, *thorough*, *polite*), whereas conservatives (vs. liberals) tend to be denied human-nature characteristics related to warmth (e.g., *fun-loving*,

**Table 1.** Final Participant Demographics (After Exclusions) From Study 1.

| Demographics | | Phase 1 Conservatives | Phase 1 Liberals | Phase 2 Participants |
|---|---|---|---|---|
| Age | M | 44.52 | 39.11 | 38.07 |
| | SD | 11.96 | 13.43 | 11.09 |
| | Range | 21–70 | 18–73 | 21–80 |
| Gender | Male | 52.0% | 47.9% | 56.3% |
| | Female | 48.0% | 51.0% | 43.2% |
| | Other | 0.0% | 1.0% | 1.0% |
| Race | White | 83.7% | 75.0% | 75.5% |
| | Asian | 8.2% | 14.6% | 11.5% |
| | Black | 4.1% | 3.1% | 6.8% |
| | Latinx | 33.1% | 6.3% | 4.2% |
| | Native | 0.0% | 0.0% | 1.0% |
| | Other | 1.0% | 1.0% | 1.0% |
| Ideology | M | 7.82 | 1.41 | 3.89 |
| | SD | 1.28 | 1.18 | 2.79 |
| | Range | 6–10 | 0–4 | 0–10 |
| Education | Bachelor's or higher | 64.3% | 61.5% | 62.0% |

*Note.* All participants were U.S. citizens. *Ideology* was measured on a scale from 0 = *extremely liberal* to 10 = *extremely conservative*. After exclusions, Phase 1 included $n = 98$ conservatives and $n = 96$ liberals; Phase 2 included $n = 192$ participants who were not selected on the basis of their political ideology.

*sociable*; Crawford et al., 2013). Study 1 examined whether patterns like these emerge when dehumanization is measured unobtrusively rather than explicitly.

## Method

Study 1 occurred in two phases. In Phase 1, participants (half of whom were politically conservative and half of whom were politically liberal) called to mind their mental representations of either *conservatives* or *liberals*, by random assignment, while completing a 300-trial reverse-correlation experiment. In Phase 2, the composite images that were generated in Phase 1—which approximate participants' mental representations of liberals and conservatives, respectively—were rated on how dehumanizing they appear.

*Participants.* Recruitment for Phases 1 and 2 was conducted via CloudResearch.com (Litman et al., 2017). Participants were adults living in the United States with a track record of high-quality survey responding (according to CloudResearch's Approved Participant List). In addition, half of the participants in Phase 1 were required to have a track record of identifying either as "conservative" or as "very conservative," whereas the other half of participants in Phase 1 were required to have a track record of identifying either as "liberal" or as "very liberal." Demographics for all final participants from Phases 1 and 2 can be found in Table 1.

Phase 1 had an a priori goal of obtaining $N = 200$ people in total ($n = 100$ conservatives; $n = 100$ liberals).[4] A total of $N = 202$ people completed Phase 1, of whom we excluded $n = 2$ (0.99%) for failing to respond "yes" to the question, "Did you take this survey seriously?" An additional $n = 6$

participants (2.97%) were excluded for being neither conservative nor liberal when asked about it directly at the end of our survey (i.e., they identified as a 5 on an 11-point scale from 0 = *extremely liberal* to 10 = *extremely conservative*). Phase 2 also had an a priori goal of obtaining $N = 200$ people. A total of $N = 200$ people completed Phase 2, of whom $n = 8$ (4.00%) were excluded for failing to respond "yes" to the question, "Did you take this survey seriously?"

### Procedure

*Phase 1.* Participants in Phase 1 completed a standard reverse-correlation procedure that was designed to approximate how they mentally represented political ingroup members versus political outgroup members. In particular, participants viewed 300 pairs of blurry faces, and their task was to choose the face in each pair that looked more like a conservative or that looked more like a liberal (by random assignment). The face pairs themselves were presented in a randomized order for each participant, and were generated by imbuing a black-and-white base image with random visual noise (Dotsch, 2016; Dotsch & Todorov, 2012; see Figure 1). In every pair of faces, one image was created by adding random visual noise to a base image, and the other was created by adding the inverse of that noise to the same base image (see Figure 1). The base image that was used was an averaged, neutrally expressive male face (taken from the AKDEF database; see Lundqvist & Litton, 1998). The use of this particular base image is common in research that uses the reverse-correlation paradigm (e.g., Dotsch & Todorov, 2012).

After the task was over, the research team computed composite images of the faces that participants chose in each
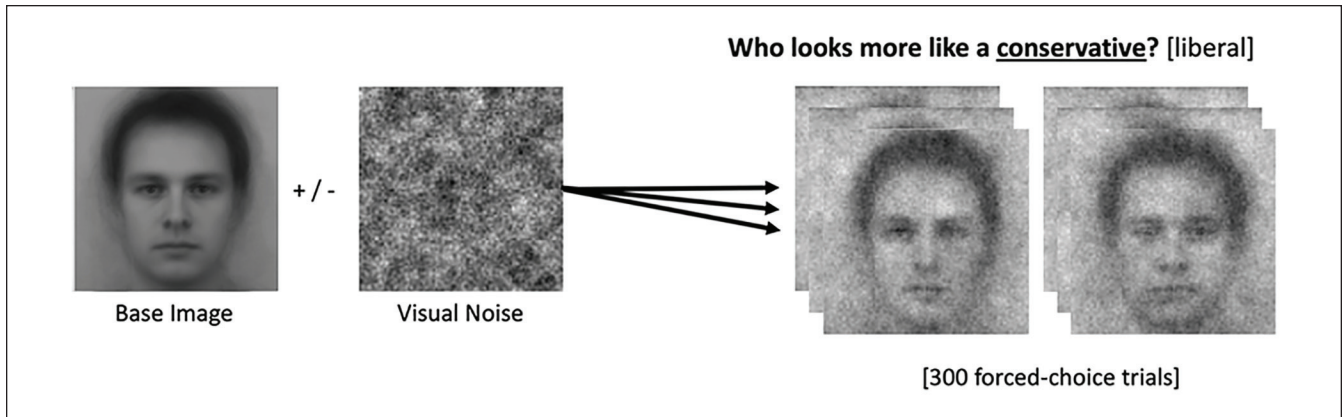
**Figure 1.** Base Image on Which Reverse Correlation Task Trials Were Based.
*Note.* One of 300 possible reverse correlation trials is depicted on the right. Each trial is a forced-choice task between two faces: one resulting from adding random visual noise to a base image, the other resulting from subtracting that same visual noise from the same base image. The base image that was used in this task, depicted on the left, is an averaged, neutrally expressive male face (taken from the AKDEF database; see Lundqvist & Litton, 1998).
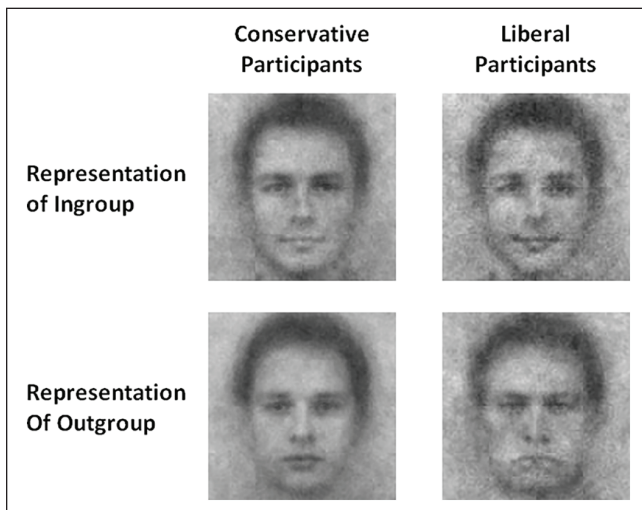


**Figure 2.** Composite Images of Political Ingroup and Outgroup Members (Study 1).
*Note.* Study 1 participants' composite images (i.e., mental representations) of political ingroup members (top row) and political outgroup members (bottom row), respectively, broken down by whether participants who generated these images were conservative (left-hand column) or liberal (right-hand column). These images were computed by averaging together participants' selections during the reverse-correlation procedure.

condition, broken down by whether the participants themselves were liberal versus conservative. To create the composite image of how conservatives mentally represent ingroup members, we created a morphed average of all the faces conservative participants chose while thinking of "a conservative." To create the composite image of how conservatives mentally represent outgroup members, we created a morphed average of all the faces conservative participants chose while thinking of "a liberal" (see Figure 2, left-hand column). Likewise, to create the composite image of how liberals mentally represent ingroup members, we created a

morphed average of all the faces liberal participants chose while thinking of "a liberal." To create the composite image of how liberals mentally represent outgroup members, we created a morphed average of all the faces liberal participants chose while thinking of "a conservative" (see Figure 2, right-hand column).

*Phase 2.* In Phase 2, new participants, who knew nothing about the composite images or where they came from, provided ratings of all four images in Figure 2. Images were rated, in a randomized order, on two main dependent variables: (a) *savagery-based* dehumanization and (b) *immaturity-based* dehumanization. For each image, participants were asked "how much does this person appear to be _____," on a scale from 1 = *not at all* to 7 = *very much*. Savagery-based dehumanization was measured with the following three items: "savage, aggressive"; "lacking morals"; and "barbaric, cold-hearted" (McDonald's ω range: .72–.91). Immaturity-based dehumanization was measured with a different set of three items: "immature, irresponsible"; "irrational, illogical"; and "naïve, gullible" (McDonald's ω range: .78–.79).[5] Items were adapted from Kteily et al. (2015) and are thus aligned more with the blatant than subtle tradition of measuring dehumanization (cf. Crawford et al., 2013).

In addition, Phase 2 participants indicated the extent to which they felt warm (vs. cold) toward each image on a standard feeling thermometer scale, as well as the extent to which each image appeared "un-evolved" according to Kteily et al.'s (2015) Ascent of Human measure of blatant dehumanization. Feeling thermometer judgments were made on a scale from 0 = *very cold* to 100 = *very warm*, whereas Ascent of Human judgments were made on a scale from 0 = *least evolved* to 100 = *most evolved* (reverse-scored such that higher numbers reflect higher levels of dehumanization). Feeling thermometer ratings will be used throughout this

**Table 2.** Standardized Relationships Between Dependent Measures in Study 1.

| Dependent variable | M (SE) | Savage | Immature | Ascent |
|---|---|---|---|---|
| Savage | 2.85 (0.07) | — | | |
| Immature | 3.28 (0.07) | .58*** | — | |
| Ascent | 19.70 (1.26) | .65*** | .49*** | — |
| Warmth | 56.55 (1.11) | −.78*** | −.35*** | −.67*** |

*Note.* Savage = savagery-based dehumanization; Immature = immaturity-based dehumanization; Ascent = Ascent of Human dehumanization; and Warmth = feeling thermometer ratings. Relationships (βs) between image ratings were computed by regressing each z-standardized dependent variable onto each z-standardized predictor in a multilevel model with one random effect: a random effect of participant intercept. This random effect adjusts for the fact that ratings of all four images were nested within person in (Phase 2 of) Study 1.
***p < .001.

article as a covariate. Findings related to the Ascent of Human measure, which replicate the patterns of dehumanization reported in this article, can be found in the online supplement. Means, standard errors, and standardized relationships between all dependent measures can be found in Table 2. Of note, this table reveals that although savagery- and immaturity-based dehumanization are positively related to one another (suggesting evidence of convergent validity: β = 0.58, p < .001), so too are they differentially related to the construct of warmth (suggesting evidence of divergent validity; βs of –0.78 and –0.35, respectively).

## Results

To analyze the data, dehumanization ratings in Study 1 were regressed, in a multilevel model, onto orthogonal contrasts that represented the 2 × 2 × 2 repeated-measures design of Phase 2's rating study. The first factor in this model was whether images had been generated by *liberals* (coded as ½) or *conservatives* (coded as –½). The second factor was whether images were of political *outgroup members* (e.g., liberals' mental representation of conservatives; coded as ½) or political *ingroup members* (e.g., liberals' mental representation of fellow liberals; coded as –½). The third and final factor was whether the type of dehumanization being measured was *savagery-based* (coded as ½) or *immaturity-based* (coded as –½). This model contained just one random effect: a random effect of participant intercept, which adjusted for the fact that the 2 × 2 × 2 design of Phase 2's rating study was nested within person. Finally, this model included (z-standardized) feeling thermometer ratings as a covariate. The reason for including this covariate was so that we could estimate outgroup dehumanization in mental representations above and beyond outgroup prejudice (i.e., the tendency to mentally represent the outgroup as more negative-looking than the ingroup). Critically, this allows for the estimation of dehumanization-specific effects. As in prior work (see Petsko et al., 2021), the inclusion of this covariate does not change the interpretation of findings reported in this article, but it does reduce the magnitude of outgroup dehumanization effects.[6] This analytic strategy and sample size gave us at least 80% power to detect main effects as small as β = 0.14,

two-way interactions as small as β = 0.21, and three-way interactions as small as β = 0.34.[7]

The research question of interest in Study 1 was whether liberals and conservatives cognitively emphasize different types of dehumanization when mentally representing one another. To examine this research question, dehumanization ratings of the mental representations obtained in Study 1 were subjected to the 2 × 2 × 2 model described above. This analysis revealed, first, a main effect of whether the mental representations were of political ingroup members versus political outgroup members. Mental representations of political outgroup members were rated as more dehumanizing (M = 3.35, SE = 0.08), on average, than mental representations of political ingroup members (M = 2.79, SE = 0.08), $M_{diff}$ = 0.55, 95% CI [0.40, 0.70], β = 0.34, F(1, 1475) = 52.30, p < .001. In addition, the magnitude of this main effect was moderated by whether the people generating the mental representations were themselves liberal versus conservative— two-way interaction: β = 0.39, F(1, 1413) = 29.92, p < .001. The nature of this interaction was such that the tendency for outgroup representations to be rated as more dehumanizing than ingroup representations was more extreme when representations had been generated by liberals, $M_{diff}$ = 0.87, 95% CI [0.64, 1.09], β = 0.54, F(1, 1480) = 58.43, p < .001, than when they had been generated by conservatives, $M_{diff}$ = 0.24, 95% CI [0.10, 0.39], β = 0.15, F(1, 1380) = 10.88, p < .001. Finally, this analysis revealed that the two-way interaction reported above interacted with dehumanization type (i.e., whether the images were being rated on how savage they appear versus how immature they appear), three-way interaction: β = 1.49, F(1, 1341) = 157.18, p < .001. This three-way interaction suggests that liberals and conservatives do not cognitively emphasize savagery-based and immaturity-based dehumanization to equal degrees when calling to mind representations of the outgroup (vs. the ingroup).

To decompose the three-way interaction between generator ideology (liberal, conservative), representation type (ingroup, outgroup), and dehumanization dimension (savage, immature) described above, we conducted spotlight analyses to examine what all model effects looked like among ratings of ingroup representations and outgroup representations,
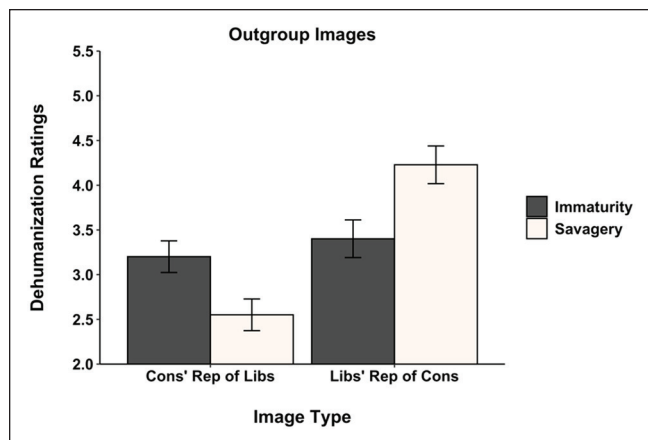
**Figure 3.** Ratings of Composite Images (of Political Outgroup Members) in Study 1.
*Note.* Cons' Rep of Libs = conservatives' representation of liberals; Libs' Rep of Cons = liberals' representation of conservatives. Means are estimated while controlling for image valence and are encompassed by 95% confidence intervals.

respectively. Examining model effects among ratings of ingroup images revealed that both liberals and conservatives tend to mentally represent themselves as less dehumanized on the dimension of savagery ($M = 2.32$, $SE = 0.08$) than on the dimension of immaturity ($M = 3.26$, $SE = 0.08$), $M_{\text{diff}} = -0.95$, 95% CI [$-1.08$, $-0.81$], $\beta = -0.59$, $F(1, 1341) = 195.66$, $p < .001$.[8] Examining model effects among outgroup images, however, told a different story. In particular, there was evidence that whether savagery or immaturity was cognitively emphasized in outgroup images depended on whether the image generators were liberal versus conservative, two-way interaction: $\beta = 0.92$, $F(1, 1341) = 119.31$, $p < .001$. Specifically, liberals mentally represented conservatives as significantly more savage-looking than immature-looking: $M_{\text{diff}} = 0.83$, 95% CI [0.63, 1.02], $\beta = 0.51$, $F(1, 1344) = 67.65$, $p < .001$. In contrast, conservatives mentally represented liberals as significantly more immature-looking than savage-looking: $M_{\text{diff}} = -0.65$, 95% CI [$-0.85$, $-0.45$], $\beta = -0.40$, $F(1, 1344) = 41.78$, $p < .001$. In other words, liberals and conservatives did indeed appear to cognitively emphasize different aspects of dehumanization when calling to mind the political outgroup (see Figure 3).[9]

### Discussion

The research question of interest in Study 1 was whether liberals and conservatives would differ in the types of dehumanization that they cognitively emphasize when mentally representing one another. Study 1 revealed that indeed, conservatives' dehumanization of liberals emphasized immaturity more than savagery and that liberals' dehumanization of conservatives emphasized savagery more than immaturity. These findings go beyond extant research on political dehumanization in that (a) these findings document political

dehumanization measured implicitly (in mental representations) rather than explicitly (on self-report), and in that (b) these findings measure dehumanization not as a unitary construct (e.g., by using the Ascent of Human measure; Landry et al., 2021; Moore-Berg et al., 2020), but at the level of specific *types* of dehumanization. This latter contribution is noteworthy, as distinct types of dehumanization are implicated in having distinct behavioral responses (e.g., different ways of expressing intergroup hostility; Kteily & Landry, 2022). To the extent that liberals and conservatives diverge in the cognitive underpinnings of their political dehumanization, they may be expected to discriminate against political outgroup members in divergent ways. We elaborate on this topic further in the general discussion.

## Study 2

Whereas the focus of Study 1 was on political dehumanization, the focus of Study 2 was on political *meta*-dehumanization. In particular, Study 2 was designed to examine whether liberals and conservatives are sensitive to how they are represented in the minds of political outgroup members. In general, existing research on political meta-dehumanization suggests that liberals and conservatives are insensitive to how much the outgroup dehumanizes them—specifically that liberals and conservatives routinely overestimate how negatively the outgroup feels (Landry et al., 2021; Lees & Cikara, 2020; Moore-Berg et al., 2020). However, these studies have all utilized self-report measures alone, and they have not examined partisans' potential sensitivity to *types* of dehumanization. In Study 2, we used reverse correlation to index partisans' *meta*-representations—that is, their representations of how they think the outgroup represents the ingroup. We then had meta-representations rated alongside actual outgroup representations (collected in Study 1), enabling a test of whether partisans can accurately project how they are represented in the minds of political outgroup members. Of note, although reverse correlation has been used to estimate representations of other people (e.g., Brown-Iannuzzi et al., 2017), or occasionally to estimate representations of oneself (e.g., Moon et al., 2020), it has never, to our knowledge, been used to estimate representations of oneself from the perspective of other people. This is to say that the use of reverse correlation to assess meta-representations is a novel contribution of Study 2.

### Method

Study 2 occurred in two phases. In Phase 1, participants (half of whom were politically conservative and half of whom were politically liberal) called to mind their *meta*-representations while completing a 300-trial reverse-correlation study. In Phase 2, composite images of meta-representations were rated alongside composite images of how the ingroup is actually represented by the political outgroup (depicted

**Table 3.** Final Participant Demographics (After Exclusions) From Study 2.

| Demographics | | Phase 1 Conservatives | Phase 1 Liberals | Phase 2 Participants |
|---|---|---|---|---|
| Age | M | 43.89 | 38.94 | 40.01 |
| | SD | 13.73 | 11.31 | 12.13 |
| | Range | 22–74 | 22–68 | 21–74 |
| Gender | Male | 50.0% | 49.0% | 54.0% |
| | Female | 50.0% | 47.0% | 44.6% |
| | Other | 0.0% | 4.0% | 0.5% |
| Race | White | 84.1% | 71.4% | 73.8% |
| | Asian | 2.3% | 10.2% | 11.4% |
| | Black | 11.4% | 6.1% | 5.0% |
| | Latinx | 0.0% | 10.2% | 5.4% |
| | Native | 0.0% | 0.0% | 0.5% |
| | Other | 0.0% | 2.0% | 2.5% |
| Ideology | M | 8.14 | 1.41 | 4.31 |
| | SD | 1.46 | 1.29 | 3.06 |
| | Range | 6–10 | 0–4 | 0–10 |
| Education | Bachelor's or higher | 45.5% | 59.2% | 56.9% |

*Note.* All participants were U.S. citizens. *Ideology* was measured on a scale from 0 = *extremely liberal* to 10 = *extremely conservative*. After exclusions, Phase 1 included *n* = 44 conservatives and *n* = 49 liberals; Phase 2 included *n* = 202 participants who were not selected on the basis of their political ideology.

previously; see Figure 2, bottom row).[10] Obtaining ratings of meta-representations in comparison with actual representations from the perspective of the political outgroup enabled us to examine the question of whether people can accurately anticipate how they are represented in the minds of others.

*Participants.* Recruitment for Phases 1 and 2 was conducted via CloudResearch.com (Litman et al., 2017). As in Study 1, participants were adults living in the United States with a track record of high-quality survey responding (according to CloudResearch's Approved Participant List). In addition, half of the participants in Phase 1 were required to have a track record of previously identifying either as "conservative" or as "very conservative"; the other half of participants in Phase 1 were required to have a track record previously identifying either as "liberal" or as "very liberal." Demographics for all final participants from Phases 1 and 2 can be found in Table 3.

Phase 1 had an a priori goal of obtaining *N* = 100 people in total (*n* = 50 conservatives; *n* = 50 liberals). A total of *N* = 100 people completed Phase 1, of whom we excluded *n* = 5 (5.00%) for failing to respond "yes" to the question, "Did you take this survey seriously?" An additional *n* = 2 participants (2.00%) were excluded for being neither conservative nor liberal when asked about it directly at the end of our survey (i.e., identifying as a 5 on an 11-point scale from 0 = *extremely liberal* to 10 = *extremely conservative*). Phase 2 had an a priori goal of obtaining *N* = 200 people. A total of *N* = 205 people completed Phase 2, of whom *n* = 3 (1.46%) were excluded for failing to respond "yes" to the question, "Did you take this survey seriously?"

*Procedure*

*Phase 1.* Participants in Phase 1 completed a 300-trial reverse-correlation task that was designed to approximate their *meta*-representations—that is, their representations of how the outgroup represents the ingroup. For each pair of faces, conservatives were asked to reflect on the question "Who would a liberal think looks more like a conservative?" In contrast, liberals were asked to reflect on the question, "Who would a conservative think looks more like a liberal?" The face pairs and base image used in Study 2 were the same as those used in Study 1.

After the task was over, the research team computed composite images of the faces that liberals and conservatives, respectively, chose while calling to mind their meta-representations. For example, to create a composite image of conservatives' meta-representation (i.e., how conservatives, on average, think liberals mentally represent conservatives), we created a morphed average of all the faces conservative participants chose during the reverse-correlation task (for conservatives' and liberals' respective meta-representations, see Figure 4, top row).

*Phase 2.* In Phase 2, new participants provided ratings of four images in a within-person rating study: composite images of liberals' and conservatives' meta-representations (Figure 4, top row), as well as composite images of how liberals and conservatives actually represent one another (Figure 4, bottom row; note that these are the same images shown previously, in the bottom row of Figure 2). All four images were rated, in a randomized order, on the same two dependent variables described in Study 1: (a) *savagery-based* dehumanization (McDonald's ω range: .84–.88), and
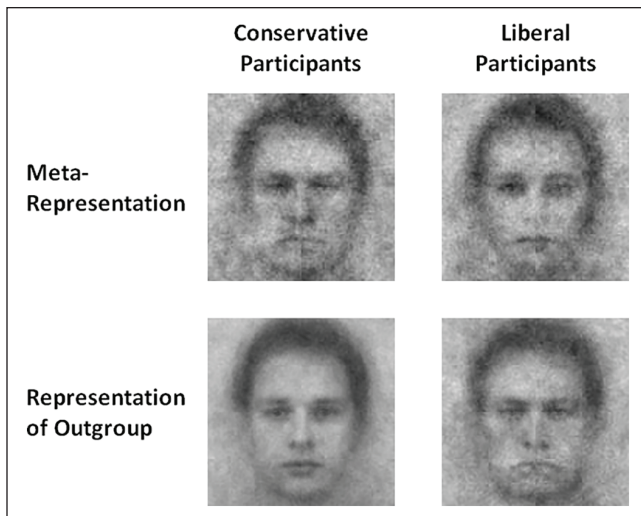
**Figure 4.** Composite Images of Meta-Representations (Study 2) Versus Outgroup Representations (Study 1).
*Note.* Composite images (i.e., mental representations) of meta-representations (top row) came from Study 2. Composite images of outgroup representations (bottom row) came from Study 1. Note that the outgroup images depicted here are the same as those shown previously. The question of interest in Study 2 is whether images that are diagonal to each other are similar (e.g., whether conservatives' meta-representation, on the top left, is similar to how liberals actually represent conservatives, on the bottom right).

(b) *immaturity-based* dehumanization (McDonald's $\omega$ range: .72–.76).

In addition, Phase 2 participants indicated the extent to which they felt warm (vs. cold) toward each image on a standard feeling thermometer scale (made on a scale from 0 = *very cold* to 100 = *very warm*), as well as the extent to which each image appeared un-evolved according to Kteily et al.'s (2015) Ascent of Human measure (0 = *least evolved* to 100 = *most evolved*, reverse-scored such that higher numbers reflect higher dehumanization). Findings related to the Ascent of Human measure, which replicate the patterns of dehumanization reported in this article, can be found in the online supplement. Means, standard errors, and standardized relationships between all dependent measures in Study 2 can be found in Table 4. Once again, we wish to note that although savagery- and immaturity-based dehumanization are positively related to one another (suggesting evidence of convergent validity: $\beta = 0.58$, $p < .001$), so too are they differentially related to the construct of warmth (suggesting evidence of divergent validity; $\beta$s of −0.78 and −0.35, respectively).

## Results

To analyze the data, dehumanization ratings in Study 2 were regressed, in a multilevel model, onto orthogonal contrasts that represented the $2 \times 2 \times 2$ repeated-measures design of Phase 2's rating study. The first factor in this model was whether images reflected representations of *conservatives* (coded as ½) or *liberals* (coded as −½). The second factor was whether images reflected *meta-representations* (e.g., liberals' representation of what conservatives think of liberals; coded as ½) or *actual outgroup representations* (e.g., conservatives' actual representation of liberals; coded as −½). The third and final factor was whether the type of dehumanization being measured was *savagery-based* (coded as ½) or *immaturity-based* (coded as −½). This model contained just one random effect: a random effect of participant intercept, which adjusted for the fact that the $2 \times 2 \times 2$ design of Phase 2's rating study was nested within person. Finally, this model included (*z*-standardized) feeling thermometer ratings as a covariate. Again, the reason for including this covariate was so that we could estimate outgroup dehumanization in mental representations above and outgroup prejudice, thereby isolating dehumanization-specific effects. According to Monte Carlo simulations, this analytic strategy and sample size gave us at least 80% power to detect main effects as small as $\beta = 0.15$, two-way interactions as small as $\beta = 0.19$, and three-way interactions as small as $\beta = 0.35$.

The research question of interest in Study 2 was whether liberals and conservatives are sensitive to how they are represented in the minds of political outgroup members. To examine this research question, dehumanization ratings of the mental representations in Figure 4 were subjected to the $2 \times 2 \times 2$ model described above. This analysis revealed, first, a two-way interaction between who was being represented in the images (conservatives vs. liberals) and whether the images were of meta-representations versus actual outgroup representations, two-way interaction: $\beta = -0.32$, $F(1, 1438) = 26.90$, $p < .001$. Breaking down this interaction revealed that conservatives' meta-representation was rated as less dehumanized ($M = 3.87$, $SE = 0.08$) than liberals' actual representation of conservatives ($M = 4.14$, $SE = 0.08$): $M_{diff} = -0.27$, 95% CI [−0.40, −0.14], $\beta = -0.17$, $F(1, 1414) = 16.22$, $p < .001$. However, for liberals, this pattern was reversed. That is, liberals' meta-representation was rated as more dehumanized ($M = 3.49$, $SE = 0.08$) than conservatives' actual representation of liberals ($M = 3.26$, $SE = 0.08$): $M_{diff} = 0.23$, 95% CI [0.10, 0.37], $\beta = 0.15$, $F(1, 1440) = 11.63$, $p < .001$. Thus, in terms of total dehumanization (collapsing across savagery and immaturity), conservatives and liberals were both inaccurate at projecting how dehumanized they would be in the minds of political outgroup members. But whereas liberals overestimated how much they were dehumanized, conservatives underestimated how much they were dehumanized.

The findings above suggest that both liberals and conservatives may be inaccurate about the overall extent to which they are dehumanized in the minds of political outgroup members (i.e., the absolute magnitude of dehumanization). But are they at all accurate about the *relative dimensions* along which they are dehumanized? Maybe so. The same $2 \times 2 \times 2$ analysis described above also provided evidence of

**Table 4.** Standardized Relationships Between Dependent Measures in Study 2.

| Dependent variable | M (SE) | Savage | Immature | Ascent |
|---|---|---|---|---|
| Savage | 3.72 (0.06) | — | | |
| Immature | 3.66 (0.06) | .49*** | — | |
| Ascent | 25.50 (1.40) | .52*** | .37*** | — |
| Warmth | 43.18 (1.09) | −.78*** | −.34*** | −.64*** |

*Note.* Savage = savagery-based dehumanization; Immature = immaturity-based dehumanization; Ascent = Ascent of Human dehumanization; and Warmth = feeling thermometer ratings. Relationships (βs) between image ratings were computed by regressing each z-standardized dependent variable onto each z-standardized predictor in a multilevel model with one random effect: a random effect of participant intercept. This random effect adjusts for the fact that ratings of all four images were nested within person in (Phase 2 of) Study 2.
***$p < .001$.
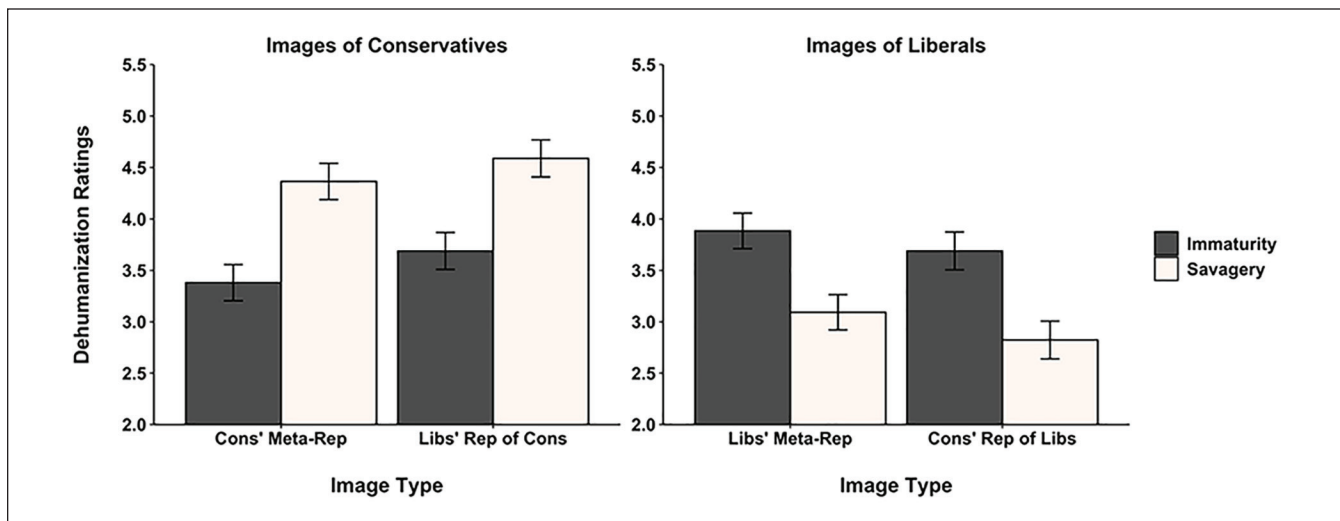


**Figure 5.** Ratings of Composite Images (Meta- vs. Outgroup-CIs) in Study 2.
*Note.* Ratings of "conservative" images are depicted on the left; ratings of "liberal" images are depicted on the right. CI = confidence interval; Cons' Meta-Rep = conservatives' meta-representation; Libs' Rep of Cons = liberals' representation of conservatives; Libs' Meta-Rep = liberals' meta-representation; Cons' Rep of Libs = conservatives' representation of liberals. Means are estimated while controlling for image valence and are encompassed by 95% CIs.

a two-way interaction between who was being represented in the images (conservatives vs. liberals) and dehumanization dimension (savagery, immaturity): β = 1.14, $F(1, 1410)$ = 364.05, $p < .001$. Breaking down this interaction revealed that, similar to what was observed in Study 1, images of conservatives tended to be rated as significantly more savage-looking ($M = 4.48$, $SE = 0.08$) than immature-looking, $M = 3.53$, $SE = 0.08$): $M_{diff} = 0.94$, 95% CI [0.81, 1.07], β = 0.61, $F(1, 1410) = 205.43$, $p < .001$, whereas images of liberals tended to be rated as significantly more immature-looking ($M = 3.78$, $SE = 0.08$) than savage-looking ($M = 2.96$, $SE = 0.08$): $M_{diff} = -0.83$, 95% CI [−0.70, −0.96], β = −0.54, $F(1, 1410) = 160.00$, $p < .001$. Moreover, this two-way interaction was *unmoderated* by whether the images were of meta-representations versus actual outgroup representations, three-way interaction: β = 0.01, $F(1, 1410) < 0.01$, $p = .96$. What this suggests is that the relative amount by which savagery versus immaturity was featured in representations held equally regardless of whether the images were of how one *expected* to be represented by the political

outgroup, or whether they were instead of how one was actually represented by the political outgroup. In other words, the evidence suggests although liberals and conservatives may be inaccurate about the overall amount by which they are dehumanized, they appear to have a sense of *how* they are dehumanized (see Figure 5): Liberals correctly perceive that they are primarily dehumanized by conservatives as immature (vs. savage), whereas conservatives correctly perceive that they are primarily dehumanized by liberals as savage (vs. immature).[11]

## Discussion

The focus of Study 2 was on the question of whether liberals and conservatives are sensitive to how they are represented in the minds of political outgroup members. The findings of Study 2 suggest that in general, liberals and conservatives may be insensitive to how much they are dehumanized by political outgroup members: conservatives' meta-representations underestimated actual outgroup dehumanization, whereas liberals'

meta-representations overestimated actual outgroup dehumanization. However, these findings also suggest that liberals and conservatives may be sensitive to *how* they are dehumanized: liberals' meta-representations accurately captured the relative amount by which conservatives think of liberals as more "immature" than "savage," whereas conservatives' meta-representations accurately captured the relative amount by which liberals think of conservatives as more "savage" than "immature." Thus, although the extant research literature characterizes liberals and conservatives as meta-perceptually inaccurate (e.g., Lees & Cikara, 2020; Moore-Berg et al., 2020), our findings suggest that some degree of accuracy might exist—at least with respect to the *types* of dehumanization that emerge in partisan mental representations.

## General Discussion

The reported studies were designed to weigh in on questions related to political dehumanization (Study 1) and political meta-dehumanization (Study 2), respectively. Rather than relying on self-report measures of dehumanization, which dominate the research literature on political (meta-)dehumanization, the present studies relied on an indirect method of assessing dehumanization: namely, reverse-correlation image classification (Dotsch & Todorov, 2012). The reverse-correlation task revealed, first, that liberals and conservatives both represented each other in dehumanizing ways, but that they differed in the dimensions of dehumanization that they cognitively emphasized. Whereas liberals' dehumanization of conservatives emphasized savagery more than immaturity, conservatives' dehumanization of liberals more strongly emphasized immaturity (vs. savagery). Second, the reverse-correlation task revealed that liberals and conservatives may have some insight into how they are represented in the minds of political outgroup members. Specifically, liberals' and conservatives' meta-representations appeared to accurately capture the relative amounts by which their respective groups are represented as savage versus immature by political outgroup members.

What are the implications of these findings? First, these findings suggest that liberals' and conservatives' dehumanization of each other may be underpinned by distinct cognitive belief systems—belief systems that emphasize savagery and immaturity, respectively. To the extent that liberals and conservatives diverge in the cognitive underpinnings of their political dehumanization, they may be expected to discriminate against each other in divergent ways (Kteily & Landry, 2022). For example, a tendency to represent liberals as immature might be correlated with discriminatory behaviors related to contempt: making derogatory remarks about liberals, decrying liberals' reactions as overly emotional, not taking liberals' perspective seriously when designing legislation, and the like. In contrast, a tendency to represent conservatives as savage might be correlated with discriminatory behaviors related to anger: verbally aggressing against conservatives, making policy recommendations that restrict conservatives' autonomy, blaming conservatives for societal dysfunction, and the like. In general, contempt and anger are mutually reinforcing emotions that correlate with separate behaviors: behaviors related to belittlement and aggression, respectively (Fischer & Roseman, 2007). Future research should investigate whether immaturity- and savagery-based dehumanization, respectively, are indeed related to these divergent intergroup emotions, and in turn, to these divergent intergroup behaviors.

A second implication of these findings is that liberals and conservatives may not be as meta-perceptually oblivious as previously thought (Lees & Cikara, 2020; Moore-Berg et al., 2020). Instead, it may be the case that liberals and conservatives have a sense of *how* they will be dehumanized—even if they are often oblivious to *how much* they will be dehumanized (Moore-Berg et al., 2020). Future research should investigate what elements of meta-perceptions are inaccurate versus what elements are accurate. More deeply understanding which meta-perceptual beliefs are accurate versus inaccurate can aid researchers in knowing which beliefs to target—and which beliefs not to target—when creating interventions that aim to reduce animus between liberals and conservatives (Druckman et al., 2022).[12]

## *Open Questions for Future Research*

An open question that these findings raise concerns how liberals and conservatives might respond to knowing how they are mentally represented by political outgroup members. According to past research, when partisans learn that they overestimate how dehumanized they are in the minds of political outgroup members, they themselves come to dehumanize the outgroup less (Landry et al., 2022). For example, when liberals learn that they overestimate how much conservatives dehumanize liberals, liberals themselves come to dehumanize conservatives to a lesser degree. A fruitful future direction for research on correcting meta-perceptions might be to investigate how liberals respond to learning both that they (a) overestimate how dehumanized they are in the minds of outgroup members, but that they (b) are nevertheless correct about the ways in which they are dehumanized (i.e., that they are indeed dehumanized along the dimension of immaturity more than savagery). On one hand, a two-sided message like this may cause liberals to regard this information as more credible than a message that emphasizes only the former, which could make that information more persuasive (Xu & Petty, 2022). On the other hand, a two-sided message like this may validate liberals' beliefs that their meta-perceptions are accurate, which may undercut the persuasiveness of learning that they overestimate how dehumanized they are in the minds of outgroup members. Future research should examine which of these two possibilities holds up to scrutiny, as these possibilities have opposing implications for reducing animus between liberals and conservatives.

A second open question concerns how stable mental representations, as indexed by reverse-correlation image classification, may be across time. Studies 1 and 2 were based on two parallel studies, called Studies S1 and S2, that were conducted back in 2017 (see the online supplement for complete details). A comparison between Studies 1 and 2, on one hand, and Studies S1 and S2, on the other hand, is worth briefly considering in relation to the issue of representation stability. First, it is worth noting that the key findings highlighted in this article appear to have remained stable over time. Back in 2017, liberals and conservatives significantly differed in the extent to which they cognitively emphasized savagery versus immaturity when dehumanizing one another (though notably, the interaction pattern is stronger here than it was back in 2017), and liberals and conservatives were likewise meta-perceptually accurate with respect to *how* they were dehumanized by political outgroup members. However, two findings in particular appear to have been more volatile over time. Although in Study 1 we found that liberals dehumanized conservatives more strongly than conservatives dehumanized liberals, back in 2017, we found that liberals and conservatives dehumanized each other to equivalent degrees (see Study S1). Likewise, although in Study 2 we found that conservatives underestimated the total amount by which they were dehumanized by liberals, we found back in 2017 that conservatives were accurate at projecting the total amount by which they were dehumanized by liberals (see Study S2). The take-home conclusion that we wish to leave readers with—in light of these patterns of convergence and divergence—is that partisans' mental representations appear to be stable with respect to the *types* of dehumanization that they capture, but more variable with respect to the *total levels* of dehumanization that they capture. Future work ought to be dedicated to examining issues of representation stability and change more directly, as research on this topic is nascent yet theoretically consequential.

### Limitations of the Present Work

A limitation of the present analysis is that it cannot be used to specify what causal relations exist between dehumanization in partisans' mental representations, on one hand, and partisans' expressions of intergroup animosity, on the other hand. For example, in the preceding paragraph, we suggested that total levels of partisan dehumanization, as indexed by mental representations, may fluctuate with time. Is it the case that fluctuations in intergroup animosity play a causal role in shaping fluctuations in partisans' mental representations? Or is it instead the case that fluctuations in partisans' mental representations play a causal role in shaping the levels of intergroup animosity they express? Ultimately, the present analysis cannot adjudicate between these possibilities, although our suspicion is that mental representations may be both caused by, and causally predictive of, intergroup animus. That is, dehumanization in one's mental representations may be thought of as

indexing partisan animosity that exists within a culture (for a similar argument, see: Payne et al., 2017), as well as driving partisans' inclinations to express that intergroup animosity. More research is required on this particular point, however. To our knowledge, causal connections between one's mental representations and one's downstream behaviors, although theoretically reasonable, have not yet been established.

A second limitation of the present analysis is that even if implicit dehumanization *does* predict downstream behavior—as we have suggested it may—it remains unclear whether implicit dehumanization can predict downstream behavior above and beyond what can be captured from explicit measures. Generally speaking, implicit attitudes have been shown to predict attitude-relevant behaviors above and beyond explicit attitudes (see Kurdi et al., 2019, for a recent meta-analysis). However, measurement of implicit dehumanization using reverse correlation specifically, rather than of more general implicit attitudes, has not to our knowledge been examined as a predictor of intergroup behavior. Thus, while it may be the case that the tendency to implicitly represent a social group as immature is predictive of (contempt-related) patronizing behavior, or that the tendency to implicitly represent a social group as savage is predictive of (anger-related) aggressive behavior—perhaps over and above what one explicitly states about a social group—this has not yet been validated. Future work should dedicate time and resources to investigating whether dehumanization that is indexed with reverse correlation can indeed explain variance in behavioral outcomes above and beyond explicit dehumanization. Such an investigation would be informative about the necessity of relying on implicit measures in addition to explicit measures when examining partisan dehumanization.

A third limitation of the present analysis is that it relies on reverse correlation, which is a somewhat recent psychological instrument whose properties are not fully understood (Cone et al., 2021). As such, it remains unclear to what extent properties of the instrument, or even properties of how people engage with the instrument, might have influenced the results reported here. For example, one property of the instrument that can influence corresponding results is which base image is used during the reverse-correlation task (Dotsch & Todorov, 2012). It is possible that data patterns reported here might have looked somewhat different if we had used an androgynous or racially ambiguous base image (e.g., Gallagher & Bodenhausen, 2021) rather than the White-male base image that we employed. Future research should examine to what extent partisan representations are shaped by the base image that serves as the foundation for those representations. A separate issue concerns how it is that people engage with the instrument. For example, although we prompted participants to choose whom they, personally, thought looked more like "a conservative" or "a liberal," it is conceivable that they used cultural stereotypes to guide their image selections rather than their own personal beliefs. If cultural stereotypes were what was guiding participants' selections—for example, stereotypes related to liberals

seeming young or conservatives seeming masculine (e.g., Koch, 2000; Rothschild et al., 2019)—it could have been these stereotypes, rather than participants' personally held views per se, that led to the observed emphases on immaturity and savagery, respectively. Future research should be dedicated to understanding whether reverse-correlation tasks measure participants' idiosyncratic representations, as is commonly argued, or whether they instead measure participants' impressions of cultural stereotypes. Such a distinction may have implications for how strongly individuals' representations can be expected to correlate with individuals' idiosyncratic behaviors.

## Concluding Remarks

Political polarization is a quickly growing problem in the United States (Finkel et al., 2020). This problem is so severe that liberals and conservatives report feeling blatantly dehumanized by those across the aisle. This feeling, in turn, predicts reciprocal dehumanization of political outgroup members, and it predicts a willingness to subvert democratic norms out of spite for those across the aisle (Moore-Berg et al., 2020). If social scientists wish to prevent democratic norms from eroding in the United States, they need to better understand the psychological bases of political dehumanization and meta-dehumanization, respectively. The present studies were an attempt to do precisely that. Findings suggest that liberals and conservatives dehumanize each other along divergent dimensions, and that liberals and conservatives may be sensitive to the dimensions along which they are dehumanized. Thus, political dehumanization may not be as monolithic as previously thought, and liberals and conservatives may not be as meta-perceptually oblivious as previously thought. Our hope is that these insights can be leveraged to improve intergroup relations among liberals and conservatives—or at least, to prevent these relations from deteriorating further.

### Data Availability Statement

All data, survey materials, and R scripts, and pre-registration documentation associated with this project are available on the Open Science Framework website: https://osf.io/483v7/.

### ORCID iD

Christopher D. Petsko [iD] https://orcid.org/0000-0002-8795-0003

### Supplemental Material

Supplemental material is available online with this article.

## Notes

1. Of note, there are other partisan stereotypes that also accord with the divergence we are championing. For example, liberals are often stereotyped as youthful and feminine, whereas conservatives tend to be stereotyped as more mature and masculine (Ahler & Sood, 2018; Koch, 2000; Rothschild et al., 2019). Partisan stereotypes related to youth and masculinity, respectively, may partly undergird those related to immaturity and savagery.

2. This condition was designed to assess a research question that is unrelated to the present article (specifically, the question of whether meta-representations are similar to how one represents one's own political group; see pp. 18–19 of the online supplement for full details).

3. The main difference between pre-registered analyses and the analyses presented here is that we pre-registered an intention to treat savagery- and immaturity-based dehumanization as separate dependent variables; however, here we opted instead to treat these as a *within-person* factor in all analyses. Treating these as a within-person factor enabled statistical tests of whether savagery is featured significantly more strongly than immaturity (or vice versa) in the minds of partisans. See OSF for a folder that contains pre-registration documentation, and see the online supplement for details on what results look like when using pre-registered analysis plans: https://osf.io/483v7/.

4. Previous research indicates that $n = 50$ people per condition is sufficient to arrive at stable composite images from reverse-correlation tasks (Petsko et al., 2021).

5. Faces were also rated in Study 1 and Study 2 on five additional traits for exploratory purposes (see online supplement). However, our intention was always to create savagery- and immaturity-based dehumanization from the six items described and analyzed in text (as can be confirmed by the pre-registration documentation on OSF).

6. For example, without controlling for feeling thermometer ratings, representations of political outgroup members are rated as 0.79 standard deviations more dehumanizing, on average, than mental representations of ingroup members ($p < .001$). The inclusion of feeling thermometer ratings as a covariate reduces the magnitude of this effect by 57% (down to 0.34 standard deviations).

7. All power analyses were conducted using the "simr" package in R (Green & MacLeod, 2016), which runs Monte Carlo simulations on one's models to arrive at power estimates (for more on this technique, see Bolger et al., 2012). Standard betas (βs) were computed by *z*-standardizing each dependent variable associated with each model from this article. Because all models contained orthogonal contrasts that were centered on zero and had a range of one, standard betas for main effect tests can be interpreted similarly to Cohen's *d*s.

8. In addition, this tendency was significantly more pronounced among liberals ($β = –0.87, p < .001$) than it was among conservatives ($β = –0.30, p < .001$), two-way interaction: $β = –0.58, p < .001$.

9. We wish to note that the tendency for liberals to cognitively emphasize savagery (vs. immaturity) to a greater extent than conservatives was also supported by Study S1 (i.e., the supplemental study on which Study 1 was based), although the

interaction pattern was stronger here than in Study S1. In addition, although in Study 1 we found that liberals dehumanized conservatives more strongly than conservatives dehumanized liberals, in Study S1, we found that liberals and conservatives did not differ in the extents to which they dehumanized one another. See pp. 21–23 of the online supplement for a full report.

10. As noted previously, Study 2's Phase 2 contained an additional condition that is reported in the online supplement rather than in the main text. Specifically, Phase 2 contained an additional condition in which composite images of meta-representations were rated in comparison with composite images of how the ingroup mentally represents themselves (i.e., the images in Figure 2, top row; see pp. 18–19 of the online supplement for more detail).

11. We wish to note that the tendency for liberals and conservatives to be accurate about *how* they are dehumanized—about the relative degree to which the outgroup represents the ingroup as savage versus immature—was also supported by Study S2 (i.e., the supplemental study on which Study 2 was based). However, Study S2 was more mixed as to whether liberals and conservatives are inaccurate about the overall extent to which they are dehumanized by the outgroup. See pp. 23–25 of the online supplement for a complete report on this issue.

12. For example, if conservatives underestimate how dehumanized they are in the minds of liberals, it may not be advisable for interventions to target conservatives' meta-dehumanization beliefs. Instead, interventions that undercut conservatives' beliefs that liberals are immature might be more effective at reducing political animosity.

## References

Ahler, D. J., & Sood, G. (2018). The parties in our heads: Misperceptions about party composition and their consequences. *Journal of Politics*, *80*, 964–981.

Andrighetto, L., Baldissarri, C., Lattanzio, S., Loughnan, S., & Volpato, C. (2014). Human-itarian aid? Two forms of dehumanization and willingness to help after natural disasters. *British Journal of Social Psychology*, *53*, 573–584.

Bolger, N., Stadler, G., & Laurenceau, J. P. (2012). Power analysis for intensive longitudinal studies. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 285–301). Guilford Press.

Boxell, L., Gentzkow, M., & Shapiro, J. M. (2023). Cross-country trends in affective polarization. *The Review of Economics and Statistics*. Advance online publication. https://doi.org/10.1162/rest_a_01160

Brown-Iannuzzi, J. L., Dotsch, R., Cooley, E., & Payne, B. K. (2017). The relationship between mental representations of welfare recipients and attitudes toward welfare. *Psychological Science*, *28*, 92–103.

Cassese, E. C. (2021). Partisan dehumanization in American politics. *Political Behavior*, *43*, 29–50.

Cone, J., Brown-Iannuzzi, J. L., Lei, R., & Dotsch, R. (2021). Type I error is inflated in the two-phase reverse correlation procedure. *Social Psychological and Personality Science*, *12*, 760–768.

Crawford, J. T., Modri, S. A., & Motyl, M. (2013). Bleeding-heart liberals and hard-hearted conservatives: Subtle political dehumanization through differential attributions of human nature

and human uniqueness traits. *Journal of Social and Political Psychology*, *1*, 86–104.

Dotsch, R. (2016). *Rcicr: Reverse-correlation image-classification toolbox* (R package version 0.3.4.1). https://CRAN.R-project.org/package=rcicr

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, *3*, 562–571.

Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M., & Ryan, B. (2022). (Mis)estimating affective polarization. *Journal of Politics*, *84*, 1106–1117.

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, A., & Druckman, J. N. (2020). Political sectarianism in America: A poisonous cocktail of othering, aversion, and moralization poses a threat to democracy. *Science*, *370*, 533–536.

Fischer, A. H., & Roseman, I. J. (2007). Beat them or ban them: The characteristics and social functions of anger and contempt. *Journal of Personality and Social Psychology*, *93*, 103–115.

Gallagher, N. M., & Bodenhausen, G. V. (2021). Gender essentialism and the mental representation of transgender women and men: A multimethod investigation of stereotype content. *Social Cognition*, *217*, 104887.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*, 493–498.

Hartman, R., Hester, N., & Gray, K. (2023). People see political opponents as more stupid than evil. *Personality and Social Psychology Bulletin*. Advance online publication. https://doi.org/10.1177/01461672221089451

Haslam, N., & Loughnan, S. (2014). Dehumanization and infrahumanization. *Annual Review of Psychology*, *65*, 399–423.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization. *Annual Review of Political Science*, *22*, 129–146.

Koch, J. W. (2000). Do citizens apply gender stereotypes to infer candidates' ideological orientations? *The Journal of Politics*, *62*, 414–429.

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*(5), 901–931.

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.

Kteily, N. S., & Landry, A. P. (2022). Dehumanization: Trends, insights, and challenges. *Trends in Cognitive Sciences*, *26*, 222–240.

Kunst, J. R., Kteily, N., & Thomsen, L. (2018). "You little creep": Evidence of blatant dehumanization of short groups. *Social Psychological and Personality Science*, *10*, 160–171.

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., Tomezsko, D., Greenwald, A. G., & Banaji, M. R. (2019). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *American Psychologist*, *74*, 569–585.

Landry, A. P., Ihm, E., Kwit, S., & Schooler, J. W. (2021). Metadehumanization erodes democratic norms during the 2020 presidential election. *Analysis of Social Issues and Public Policy*, *21*, 51–63. https://doi.org/10.1111/asap.12253

Landry, A. P., Schooler, J. W., Willer, R., & Seli, P. (2022). Reducing explicit blatant dehumanization by correcting exaggerated meta-perceptions. *Social Psychological and Personality Science*, *14*, 407–418. https://doi.org/10.1177/19485506221099146

Lees, J., & Cikara, M. (2020). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature Human Behavior*, *4*, 279–286.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*, 433–442.

Lundqvist, D., & Litton, J. E. (1998). *The Averaged Karolinska Directed Emotional Faces—AKDEF*. https://kdef.se/index.html

Martherus, J. L., Martinez, A. G., Piff, P. K., & Thedoridis, A. G. (2021). Party animals? Extreme partisan polarization and dehumanization. *Political Behavior*, *43*, 517–540.

Moon, K., Kim, S., Kim, J., Kim, H., & Ko, Y.-g. (2020). The mirror of the mind: Visualizing mental representations of self through reverse correlation. *Frontiers in Psychology*, *11*, 1149.

Moore-Berg, S. L., Ankori-Karlinsky, L.-O., Hameiri, B., & Bruneau, E. (2020). Exaggerated meta-perceptions predict intergroup hostility between American political partisans. *Proceedings of the National Academy of Sciences*, *117*, 14864–14872.

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.

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*, 233–248.

Petsko, C. D., Lei, R. F., Kunst, J. R., Bruneau, E., & Kteily, N. (2021). Blatant dehumanization in the mind's eye: Prevalent even among those who explicitly reject it? *Journal of Experimental Psychology: General*, *150*, 1115–1131.

Rothschild, J. E., Howat, A. J., Shafranek, R. M., & Busby, E. C. (2019). Pigeonholing partisans: Stereotypes of party supporters and partisan polarization. *Political Behavior*, *41*, 423–443.

Xu, M., & Petty, R. E. (2022). Two-sided messages promote openness for morally based attitudes. *Personality and Social Psychology Bulletin*, *48*, 1151–1166.