Running head: TEMPORAL DYNAMICS

The Temporal Dynamics of the Link between Configural Face Processing and Dehumanization

Steven G. Young[1,2], Ryan E. Tracy[2], John Paul Wilson[3], Robert J. Rydell[4], & Kurt Hugenberg[4]

1. Baruch College, City University of New York
2. The Graduate Center, City University of New York
3. Montclair State University
4. Indiana University

Word Count: 4996

Authors' Note: This research was supported by National Science Foundation grants BCS-1423765 and BCS-1748761

Address correspondence to Steven Young at steven.young@baruch.cuny.edu

science

Abstract (101 words)

The human face conveys a wealth of information, including traits, states, and intentions. Just as fundamentally, the face also signals the humanity of a person. In the current research we report two experiments providing evidence that disruptions of configural face encoding affect the temporal dynamics of categorization during attempts to distinguish human from non-human faces. Specifically, the present experiments utilize mouse-tracking and find that face inversion elicits confusion amongst human and non-human categories early in the processing of human faces. This work affords the first examination of how facial inversion affects the dynamic processes underlying categorization of human and non-human faces.

Keywords: configural face processing; dehumanization; mouse-tracking

The Temporal Dynamics of the Link between Configural Face Processing and Humanness

The human face broadcasts a wealth of socially important information, including static properties like race, sex, and personality characteristics (e.g., Carre, McCormick, & Mondloch 2009; Cloutier, Mason, & Macrae, 2005), as well as situational information like behavioral intentions and emotional states (e.g., Parkinson, 2005; Wilson & Hugenberg, 2013). Successfully extracting this information relies largely on configural face processing, a feature integration process that entails encoding the spatial relations between face parts, rather than the face parts in isolation (e.g., Maurer, LeGrande, & Mondloch, 2002). Evidence for this comes from manipulations that interfere with configural encoding. The best known of these is face inversion (Valentine, 1988; Yin, 1969), a technique which disrupts the canonical eyes-above-nose-above-mouth configuration while leaving the face parts themselves intact. Inverting faces impairs identity recognition (Rhodes, Brake, Taylor, & Tan, 1989), emotion decoding (McKelvey, 1995; Young & Hugenberg, 2010), and trait inferences (Wilson, Young, Rule, & Hugenberg, 2018).

Notably, this feature integration process is typically reserved for faces; non-face stimuli are instead processed in a ~~more~~ piecemeal, feature-based manner (McKone & Robbins, 2011; Morton & Johnson, 1991). As a result, inversion effects are ~~much~~ more pronounced for faces than other ~~classes of~~ stimuli (e.g., Yin, 1969). This domain specificity raises the interesting possibility that encoding the configuration of face parts conveys information beyond identity and emotional states. Indeed, the experience of processing a face configurally may be a bottom-up signal that we are interacting with another person who possesses fundamental capacities to think, feel, and act in human ways (see Deska & Hugenberg, 2017; Fincher, Tetlock, & Morris, 2017 for recent reviews). If true, interfering with configural face encoding should lead to diminished

capacity to recognize that faces belong to fellow humans, rather than objects or non-human animals.

Recent research supports this possibility. In an initial demonstration of this configural-to-humanness link, Hugenberg and colleagues (2016) showed that face inversion disrupts the processing of others' humanness early in the perceptual stream. For example, in their first study, inverted faces failed to spontaneously activate human-related concepts (e.g., soul), but had no effect on concepts relevant to machines (e.g., computer). Similarly (and important for the present research), disrupting configural face processing also delayed perceivers' ability to categorize human faces as human, but face inversion had no influence on perceivers' speed at categorizing chimpanzee faces as animals (presumably because configural information is not used to make animal categorizations).

Interestingly, initial evidence suggests that the configural-to-humanness link occurs for both racial outgroup and racial ingroup faces (Cassidy et al., 2017; Wilson, et al., 2018). However, this preliminary research did not employ a diverse sample of participants. Indeed, the question of how well past inversion-dehumanization effects generalize across perceiver and target race is important, especially considering evidence that configural processing occurs more strongly for same-race (Michel et al., 2006; Rhodes et al., 1989) and ingroup targets (Hugenberg & Corneille, 2009).

Just as disrupting configural processing appears to disrupt a bottom-up signal of humanness, believing a target to be inhuman may influence perceivers' use of face-typical configural processing. For example, Fincher and Tetlock (2016) had participants learn about a series of individual faces, some of whom committed egregious, non-normative acts (e.g., rape), and others that behaved ~~demonstrated~~ laudablably~~e behavior~~ (e.g., charitable donations). Fincher

and Tetlock then employed measures of configural processing to demonstrate that the faces of inhumane actors are perceptually processed less like faces and more like objects. Importantly, this object-typical processing also has implications for punishing wrongdoers -- perceivers find it easier to punish perpetrators who they do not process configurally.

Whereas past research has reliably linked configural face processing with judgments of others' humanness, no research has yet investigated the temporal dynamics of this configural-to-humanness link. Across two experiments using mouse-tracking as a measure of dehumanization, we pursue just this question. Integrating Freeman and Ambady's (2011) *Dynamic Interactive Theory* of person construal with recent evidence suggesting that configural processing affects categorization, we investigated the extent to which face inversion fails to activate "humanness" when perceiving human faces, leading to difficulty in making early categorizations of human faces as human (as compared to categorizing the humanness of robot or chimpanzee faces).

The Temporal Dynamics of Categorization

Categorizing people into discrete social groups has long been considered an effortless, efficient, and even obligatory process (see Kawakami, Amodio, & Hugenberg, 2017). Formative work ~~in the social cognitive and social neuroscience literatures~~ provides evidence that category cues are extracted from faces quickly, allowing for rapid categorization along visually salient dimensions like sex and race (e.g., Clouthier, et al., 2005; Ito & Urland, 2003). Recently, Dynamic Interactive Theory (DIT; Freeman & Ambady, 2011) has emphasized the ongoing integration of bottom-up perceptual information with top-down factors that ultimately produce categorical construals of people.

This theory stipulates that the earliest stages of processing rely on relatively coarse, imprecise information that may lead to multiple possible categorizations, before settling over

time into a stable, often unitary categorization. Consider the case of determining whether a novel face is female or male, a rapidly generated categorical distinction that often dominates person construal (Quinn & Macrae, 2005). Upon ~~first~~ encountering the face, various bottom-up cues (e.g., hair length, width-to-height ratio) interact with top down cues (e.g., perceiver stereotypes) and may activate both the female and male category. As processing proceeds over time, these competing categories are resolved as more finely grained cues provide additional information (e.g., hair length, jaw prominence). Collectively, the continuous and interactive integration of this information allows a final category decision to be made (see Freeman & Ambady, 2014).

Notably, the categorization process can be measured in real-time, using a methodological development that accompanied the growth of DIT -- mouse-tracking (Freeman, Dale, & Farmer, 2011; Freeman & Ambady, 2010). This processing-tracing methodology tasks participants with using a computer mouse to categorize stimuli into discrete categories and measures behavior in a continuous manner as the mouse is moved across the computer screen toward one or another categorization decision (see Freeman, 2018 for an overview). One particularly valuable insight offered via mouse-tracking is examining category competition as the categorization decision unfolds (Hehman, Stolier, & Freeman, 2015). For example, mouse trajectories while categorizing a masculine female face as female (i.e., correct) or male (i.e., incorrect) show an early "pull" toward the male category option that is subsequently corrected, reflecting the confusion or co-activation of the "female" and "male" categories in early stages of categorization, before the percept fully resolves as female (e.g., Freeman, Ambady, Rule, & Johnson, 2008). Using this method allows researchers to examine decision dynamics during the categorization process, providing novel insight beyond reaction time data.

The Current Experiments

The current research adopted mouse-tracking techniques in an attempt to elucidate the temporal dynamics of the face inversion-dehumanization process. More specifically, in the present work we investigate how upright versus inverted human faces, as compared to upright and inverted animal and robot faces, differentially elicit activation of humanness over time in categorization tasks. Whereas extensive past research has focused on categorization under conditions of early category uncertainty (i.e., race or sex ambiguous targets), the present research attempts to extend DIT to cases of dehumanization. Of specific interest in the present work is how these dynamic processes unfold when discriminating human from non-human stimuli, including animals and machines, in order to test for evidence of both animalistic and mechanistic dehumanization (see Haslam, 2006; Haslam & Loughnan, 2014). To elaborate, withholding humanity can manifest as mechanistic dehumanization (i.e., attributed robotic characteristics, but seen as lacking in human warmth; e.g., Bain, Park, Kwok, & Haslam, 2009), or animalistic dehumanizationed (i.e., attributed animalistic characteristics, but seen as lacking in human sophistication; e.g., Goff, Eberhardt, Williams, & Jackson, 2008). Thus, in the current work we employ robot and chimpanzee faces as comparison stimuli to demonstrate early perceptual dehumanization along both dimensions.

By employing mouse-tracking we can understand how the activation of dehumanizing concepts (i.e., likening people to machines or animals) occurs early in the category activation stream. Specifically, we hypothesize that configural face processing of upright faces will provide a strong bottom-up cue of humanness, even for these commonly dehumanized groups, allowing perceivers to easily distinguish between human and non-human stimuli. Put simply, upright faces will be easily distinguished from robots and animals, despite the perceptual similarities of these stimuli (i.e., shared configural properties). However, because inversion disrupts the

bottom-up signal of humanness, this should lead to category confusion between human and robot categories or human and animal categories, depending on the relevant comparison stimuli employed in a given experiment. ~~Put simply~~Overall, when perceivers lack the signal of humanness conveyed by configural processing of human faces, the category "human" should become more difficult to resolve early in the processing stream.

Past research has found that mechanistic dehumanization is most common for Asians, (e.g., Bain et al., 2009), whereas animalistic dehumanization is often directed toward African Americans (e.g., Goff et al., 2008). Based on this empirical precedent, we presently test for early categorization using the most relevant non-human comparisons. However, past evidence also shows that inversion impairs discriminating White human faces from chimpanzee faces (Hugenberg et al., 2016; Experiment 2), indicating that stereotype consistency is not necessary for inversion to disrupt signals of humanness. Notably, this was true even though the human and non-human faces varied in coloration (e.g., darkly colored chimpanzee faces and pale human faces).

Finally, we hypothesize that these inversion effects on categorization should be observable most strongly early in the perceptual stream. That is, the absence of the configural signal will not ultimately make it impossible for perceivers to distinguish a human's face from a robot or from an animal. Instead, absence of the ~~signal of~~ humanness signal should be observable early in the perceptual stream as initial category confusion, and overcome later as additional information (e.g., knobs in place of eyes) informs the ultimate categorical decision.

To investigate these hypotheses, in Experiment 1, ~~we ask~~ participants ~~to~~ categorize a series of upright and inverted Asian and robot faces. ~~–~~Experiment 2 offers a pre-registered conceptual replication investigating the confusion of "human" and "animal" categories using

Black targets. (e.g., Apel, 2009; Costello & Hodson, 2010; Goff, et al., 2008). Collectively, these experiments allow us to investigate how the lack of a humanness signal from inverted faces may affect categorizations between "human" and both animalistic and mechanistic categories. We report all measures, manipulations, and exclusions in these studies.

Experiment 1

In Experiment 1, participants categorized a series of upright and inverted East Asian male and human-like robot faces as either "human" or "robot." Employing a mouse-tracking methodology, of interest was whether inverted faces led to stronger category confusion inn human and non-human categories in early processing. Here, we predicted that inversion would lead to dehumanization of human faces, seen in mouse trajectories pulling toward the "robot" category more strongly than in upright stimuli.

Method

Participants. 61 participants (M_{age} = 22.78; 35 Female) completed the experiment in exchange for partial course credit. The sample was diverse (27 Asian, 14 White, 10 Latina/o, 5 Black, 5 "other"). We targeted 60 participants, which would provide 94.9% power to detect an effect size of d = .35 (between a small and medium effect) according to a power analysis conducted in PANGEA that specified a fully crossed 2 x 2 repeated-measures design, setting both participants and stimuli as random factors (Judd, Westfall, & Kenny, 2016).

Procedure. After providing informed consent, participants were verbally instructed that they were completing a brief computerized experiment on "face and object processing." Participants were instructed onscreen that they would be shown faces of humans and robots and be asked to categorize them as quickly and accurately as possible as either "human" or "robot" by clicking on the appropriate category label using the computer mouse.

The robot images were found via a google image search and then standardized, grayscaled, and cropped so that all images were the same size and front-facing[1]. To select the ideal faces for the current research, the set of 59 robot images was pre-tested on Mechanical Turk. Specifically, 60 participants viewed each face in a random order and rated how "human-like" and "face-like" each robot appeared using 1 (*not at all*) to 7 (*extremely*) Likert-type scale. These two ratings were highly correlated (r = .875) so a single composite score of humanness was created. We then selected the 15 most human robot faces for use (M = 4.84, SD = .56) and created inverted version of each. To match the number of unique robot stimuli, we selected 15 young adult Asian male targets (Minear & Park, 2004). Upright and inverted versions of each face were created and shown in grayscale. All faces were sized to 200 × 300 pixels.

Following standard mouse-tracking procedure, each trial began by using the mouse to press a "start" button located on the bottom-center of the screen. Clicking start cued the presentation of a randomly selected face which replaced the start button in the bottom-center of the screen. Category labels "Human" and "Robot" were presented in the top right and left corners of the screen, respectively, and were visible during the entire task. Participants moved the mouse from the bottom-center location to the appropriate label in the top corners of the monitor as quickly as possible. After each categorization was made the start button reappeared and participants returned the cursor to the starting position before initiating the next trial. In total, participants completed 60 trials (30 human, 30 robot, with 15 upright and 15 inverted faces in each category). To familiarize participants with the mouse-tracking procedure, eight practice trials were presented prior to the face categorization task. On these trials participants categorized fruits and vegetables.

---

[1] Stimuli are available upon request from the corresponding author

Results

Of interest was whether inverting human faces led to stronger category confusion for human and robot categories. To investigate this, we first calculated Area under the Curve (AUC scores), measuring how much mouse trajectories deviated from an "ideal" straight line during each categorization decision. Prior to analyses, data were filtered to remove errors (0.22% of trials), trials with initialization times above 500ms (1.75% of remaining trials), and trials that were more than 3 SD from a participant's average response (2.93% of remaining trials). Positive AUC scores indicate attraction to a competing category during processing (i.e., the mouse trajectory pulls toward the incorrect category before corrected midstream). All analyses were conducted using R software (Version 3.4.0; R Development Core Team, 2017) and models were built using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). Model *p*-values were calculated using R's lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017) which ran the models through the Satterthwaite approximation tests for the degrees of freedom estimations while the base R package confint was used to compute 95% profile confidence intervals for the fixed and random effects.

Participants' AUC values were submitted to a series of linear mixed effects that regressed AUC values on Face Type (human = .5, robot = -.5) and Orientation (upright = .5, inverted = -.5). Random effects were specified in a series of models in a step-wise fashion to determine the best fit, with the ultimate aim of fitting a model that specified random factors for both participants and stimuli (see Judd, Westfall, & Kenny, 2012). Thus, this analysis first fitted a model with random intercepts for participants only, a model with random intercepts for both participants and stimuli, a model with random intercepts for participants and stimuli with a random slope for Orientation, and a model with random intercepts for participants and stimuli

with a random slope for Face Type. Of these models, the final was determined to provide the best fit for the data, $\chi^2(3) = 211.33$, $p < .001$, thus this was the model used for our main analyses.

In this model, the random effects structure demonstrated that the random intercept for participants was significant (*SD* = .24, 95% CI [.20, .29]), as was the random slope for Face Type (*SD* = .33, 95% CI [.27, .40]). The random intercept for stimuli did not demonstrate significance (*SD* = .02, 95% CI [0, .05]), but because this did not result in an overfitted model this term was allowed to remain. This model revealed a nonsignificant effect of Face Type, *b* = -.06, *SE* = .05, *t*(59.81) = -1.34, *p* =.19, 95% CI [-.07, .004], and a nonsignificant effect of Orientation, *b* = -.03, *SE* = .02, *t*(56.49) = -1.70, *p* = .10, 95% CI [-.15, .03]. There was a significant interaction between Face Type and Orientation, *b* = -.18, *SE* = .04, *t*(56.49) = -4.76, *p* < .001, 95% CI [-.25, -.11] (see Figure 1). As predicted, inverted human faces (*M* = .69, *SD* = 1.03) had significantly higher AUC values than upright human faces, (*M* = .57, *SD* = .92), *t*(60) = 4.18, *p* < .0001, *d* = .54, 95% CI [.17, .90], indicative of stronger category confusion between the categories human and robot for inverted (relative to upright) human faces. No significant differences were found for upright (*M* = .64, *SD* = .97) versus inverted robot faces (*M* = .69, *SD* = 1.02), *t*(60) = 1.67, *p* = .10, *d* = .21, 95% CI [-.57, .15], indicating no evidence of confusion in early category activation for non-human faces.

Notably, erroneous responses- were uncommon~~occurred on 0.2% of trials~~, indicating that category confusion effects emerge early in processing ~~stages~~ but are corrected in late-stage categorization decisions.

Discussion

Experiment 1 finds evidence that disrupting configural processing of human faces can create human/non-human category confusion in early person construal. This supports earlier

research arguing that configural face processing triggers a bottom-up cue of humanness, which otherwise assists perceivers in making category distinctions earlier in the processing stream. However, even for robots selected to be especially face-like, configural processing in the absence of featural cues of humanness does not appear to influence animacy-related judgments (see also Deska, Almaraz, & Hugenberg, 2016). Thus, it is not that any inverted face is difficult to discriminate between human and non-human, but rather that the *absence* of the humanness signal for inverted human faces disrupts early human/non-human distinctions for human faces.

Experiment 2

Whereas Experiment 1 investigated the dynamic time course of inversion-to-dehumanization link for Asian faces, Experiment 2 is designed to extend these findings to Black faces. Although past research has replicated the tendency for inversion to affect overt dehumanizing judgments of Black targets (e.g., Cassidy et al., 2017), no research has yet investigated how this effect unfolds over time in early category activation. Additionally, Experiment 2 included animal faces as control stimuli. This provides an opportunity to test whether inversion leads to early stage category confusion of not only machine-like categories, but also animal-like categories. Given that dehumanization can occur along orthogonal mechanistic and animalistic dimensions (e.g., Haslam & Loughan, 2014), it is theoretically important to test whether inversion-dehumanization effects generalize across both. Together, these modifications allow us to simultaneously examine the generalizability of the temporal dynamics of the configural-to-human link both in a new target group and with a ~~quite~~ distinct type of dehumanizing judgment.

Thus, in Experiment 2, participants completed a mouse-tracking categorization task for upright and inverted Black and chimpanzee faces. We expected to find that once again inverted

human faces would be dehumanized, as demonstrated by an early tendency for category confusions, indexed by AUC values recorded via mouse-tracking.

Method

Participants. As in Experiment 1, we targeted at least 60 participants. In actuality, 87 participants (M_{age} = 23.8; 48 Female) completed the experiment in exchange for partial course credit. The sample was diverse (36 Asian, 19 White, 18 Black, 1 Latina/o, 2 "other"). No analyses were conducted until data collection was completed.

Procedure. The procedure was identical to Experiment 1, except for the following modifications. First, the human faces used here were 20 Black males used in past research (e.g., Young, Bernstein, & Hugenberg, 2010). Second, the non-human stimuli were 20 chimpanzee faces were taken from Taubert, Qureshi, & Parr (2012; see also Young, Goldberg, Rydell, & Hugenberg, 2019). All stimuli were presented in grayscale, and sized to 200 × 300 pixels.

This experiment was preregistered on AsPredicted.org (#10194). All stimuli, procedures, exclusions, and data analyses are consistent with the pre-registration plan[2]

Results

AUC values were calculated as in Experiment 1and were again regressed on Face Orientation (upright = .5, inverted = -.5), Face Type (human = .5, chimp = -.5), and their interaction in a linear mixed effects model. Prior to analyses, data were filtered to remove errors (0.32% of trials), trials with initialization times above 500ms (2.4% of remaining trials), and

[2] The pre-registration described analyzing the data using repeated-measures ANOVA rather than the linear mixed effects model reported here. We elected to report the mixed model results in the main text as a more rigorous analysis. The results are virtually identical with the ANOVA, including the Face Type × Orientation interaction, F (1,86)=10.02, p=.002. Pairwise comparisons reveal greater AUC values for inverted than upright human faces, t (86)=3.53, p=.001, but no such difference for chimp faces, t (86)=1.19, p=.266.

trials that were more than 3 SD from a participant's average response (3.2% of remaining trials). As with Experiment 1, the random effects structure was determined in a step-wise fashion, which indicated that a model specifying random intercepts for both participants and stimuli with a random slope for Face Type provided the best fit, $\chi^2(3) = 97.67$, $p < .001$. This model revealed a nonsignificant effects for Orientation, $b = -.02$, $SE = .02$, $t(73.94) = -.92$, $p = .36$, and Face Type, $b = -.04$, $SE = .03$, $t(78.41) = -1.26$, $p = .21$, and the predicted significant Face Type × Orientation interaction, $b = -.12$, $SE = .07$, $t(73.93) = -3.21$, $p = .002$ (see Figure 2). Consistent with predictions, participants had higher AUC values when classifying inverted ($M = .68$, $SD = 1.08$) versus upright ($M = .56$, $SD = .91$) human faces, $t(86) = 3.53$, $p < .0001$, $d = .38$, 95% CI [.07, .68]. No such differences were found for inverted ($M = .66$, $SD = 1.03$) versus upright chimpanzee faces ($M = .69$, $SD = 1.04$), $t(86) = 1.12$, $p = .26$, $d = .12$, 95% CI [-.18, .49]. Errors occurred- rarely~~on only 0.3% of trials~~, again indicating that the effects are due to early-stage category confusions that is resolved later in the categorization process.

Discussion

Experiment 2 once again offers evidence for dehumanization in the early stages of processing when configural encoding of human faces is disrupted via inversion. Specifically, using Black faces, we found that categorization trajectories are relatively linear for upright faces but show attraction to the non-human category when inverted. This suggests that inversion disrupts the signal of humanness typically afforded by configural processing. However, as in Experiment 1, inversion effects were asymmetrical; they did not lead to the confusion of human and animal concepts for inverted animal faces. Thus, it was not merely that inverted faces are difficult to discriminate, but that human faces specifically lacked sufficient signals of humanness to make easy early-stage categorization decisions.

General Discussion

Human faces broadcast a wealth of information, including traits and states that facilitate interpersonal functioning (Wilson & Hugenberg, 2013). The capacity to extract this information relies largely on configural encoding, a feature integration process typically reserved for faces (Valentine, 1988; Yin, 1969). The current work adds to a growing body of research demonstrating that configural encoding is a ~~fundamental~~ perceptual signal of humanity, cueing perceivers that they are interacting with another person with the range of cognitive and experiential capacities typical of humans (Deska & Hugenberg, 2018; Fincher & Tetlock, 2016; Hugenberg et al., 2016). Consistent with these recent findings, we find that disrupting configural encoding of human (but not of non-human) faces elicits human/non-human category confusions early in person construal.

Not only are these findings conceptually consistent with past research investigating the effects of configural processing on humanness judgments, but mouse-tracking affords novel advances over past research. Specifically, we see the tendency for initial categorization of inverted human faces to be systematically pulled toward a non-human categorization alternative (i.e., animals and robots, respectively). This movement toward the non-human category is clear evidence for *dehumanization* – inverted human faces lack sufficient signal of humanness to make early-stage categorizations.

The use of both animal and robot comparisons extends past research as well. In broad strokes, prominent theories of dehumanization and mind perception suggest that orthogonal dimensions of cognitive sophistication and experiential depth distinguish humans from animals and machines (Haslam & Loughnan, 2014; see also Gray, Gray, & Wegner, 2007). As a consequence, dehumanization often unfolds along either mechanistic or animalistic lines.

Whereas past research has shown that inversion can influence categorization in human versus animal tasks where the human group (White males) are not stereotypically likened to animals, presently we find that inversion can implicate both mechanistic and animalistic dimensions of dehumanization, and that this occurs for targets from non-White social groups.

Configural Encoding and Race

An additional noteworthy contribution of the present work is the inclusion of non-White targets, and a diverse sample of participants. Whereas some past research has investigated target race as a potential factor in the configural-to-humanness link (Cassidy et al., 2017), no research has yet done so using a diverse sample of participants. Collectively, the pattern of results for human faces are indistinguishable when Asian and Black faces are used. Although we do not directly test for participant race effects (we did not have adequate power to do so), we nevertheless detect robust effects of inversion on human categorizations despite participant-level variance in race. Notably, although interactions between stimulus and participant race have been observed for some inversion effects (Rhodes, et al., 1989), this does not appear to be the case for configural-to-humanness links. For example, Wilson and colleagues (2018) show that inversion affects the extraction and ascription of facial trustworthiness in White and Black faces alike (with a primarily White sample). Interestingly, inversion has even been found to have a larger impact on other-race than same-race faces when judged on trustworthiness and perceived homogeneity (Cassidy et al., 2017), perhaps reflecting the tendency for already dehumanized groups (e.g., Asians, Blacks) to co-activate human and non-human categories in the absence of the otherwise humanizing signal provided by configural processing.

The apparent generalizability of the configural-to-humanness link across target and participants race suggests that it is robust and not dependent on stereotype-consistency or

accessibility. However, beyond exhibiting the breadth of inversion effects on dehumanization, this cross-race generalizability has interesting implications for the larger configural encoding literature. For example, although some have documented links between configural processing and race or group membership (e.g., Rhodes et al., 1989), contrary evidence suggests that perceivers are capable of configurally processing upright other-race faces even in the absence of frequent contact or perceptual training (Cassidy, Boutsen, Humphreys, & Quinn, 2014; ~~Civile, Colvin, Siddiqui, & Sukhvinder, in press;~~ Weiss, Stahl, & Schweinberger, 2009; Zhao, Hayward, & Bulthoff, 2014). In a conceptually similar way, aspects of the current experimental context may have encouraged configural encoding across target races by manipulating the human versus non-human stimuli, rather than manipulating race within task (~~see~~ Young, Hugenberg, Bernstein, & Sacco, 2009). Holding race constant within each of the current experiments may have allowed for more face-specific, configural processing. Although the present work was not designed to specifically test whether the presence or absence of intergroup contexts influences dehumanization, future research may ~~well~~ benefit from doing so.

Future Directions

Several other future direction remain as well. First, perhaps inducing situations that lead to dehumanization (e.g., intergroup competition, e.g., Fiske, 2013) will make intergroup distinctions, like race, more relevant. Under these circumstances, we may see that even upright outgroup faces are subject to dehumanization. Second, whereas past research has focused on how inversion disrupts the ascription of humanness to White targets (e.g., Fincher & Tetlock, 2016; Hugenberg et al., 2016), the present research focused on ~~the non-human categories for~~ commonly dehumanized Asian and Black targets. However, the current experiments do not compare the effects of inversion for majority group (White) and minority group (Asian, Black)

targets. This was intentional, to ensure that the experimental context did not strongly trigger an intergroup context (see Young et al., 2009) and the comparison stimuli (robots and animals, respectively) were selected for their theoretical relevance for mechanistic and animalistic dehumanization. However, we do not believe the present findings are specific to these experimental conditions. Theoretically, inversion should mute the "humanness" signal normally conveyed by configural face processing, leading to dehumanization in broad and generalizable ways, such that a White face and chimp (or robot) categorization task would replicate the current results (e.g., Hugenberg et al., 2016 Experiment 2).

Finally, a pressing question is how early-stage perceptual dehumanization relates to downstream outcomes. Although past work has linked disruptions of configural encoding to negative treatment (Fincher & Tetlock, 2016), it would nevertheless be interesting to know how early-stage processes captured via mouse-tracking relate to consequential social outcomes. Indeed, other research has reliably shown that early activation as indexed by mouse-tracking leads to impactful real-world behaviors (e.g., Hehman, Carpinella, Johnson, Leitner, & Freeman, 2014). As such, downstream effects of dehumanization beyond those already known (Fincher & Tetlock, 2016) seem likely and await discovery.

Conclusion

In two experiments, we extend the logic and findings of the Dynamic Interactive Theory (Freeman & Ambady, 2011) to the early perceptual cues that perceivers use in judging others' humanity. Specifically, the current research indicates that face inversion can undermine a critical cue to others' humanity, causing human/non-human category confusion in early person construal. These findings add to a growing literature on perceptual dehumanization and lend additional support to the view that configural face processing is a bottom-up visual signal of

humanness (e.g., Deska & Hugenberg, 2018; Fincher, et al., 2017). Further, we find this early human/non-human category confusion can occur in both mechanistic and animalistic ways (e.g., Haslam & Loughnan, 2014)~~, extending past research as well~~. Finally, we also offer additional evidence that the inversion-dehumanization link generalizes across target and participant race (e.g. Cassidy et al., 2018; Wilson, et al., 2018). Collectively, the present work shows that dehumanization can have roots in perceptual and visual processing of faces, expanding the study of this pernicious effect and underscoring the social importance of understanding ~~the~~ how the early stages of the categorization process influence person construal.

Open Practices

Our preregistration plan can be accessed here: http://aspredicted.org/blind.php?x=kb4d6d.

Data accompanying these experiments are available via Mendeley.

References

Bain, P., Park, J., Kwok, C., & Haslam, N. (2009). Attributing human uniqueness and human nature to cultural groups: Distinct forms of subtle dehumanization. *Group Processes & Intergroup Relations, 12*, 789-805.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using 'Eigen' and S4 (Version 1.1-7) [Computer software]. Retrieved from http://cran.r-project.org/package=lme4

Carré J.M., McCormick C.M. & Mondloch C.J. (2009). Face structure is reliable cue of aggressive behavior. *Psychological Science, 10*, 1194-1198

Cassidy, B. S., Krendl, A. C., Stanko, K. A., Rydell, R. J., Young, S. G., & Hugenberg, K. (2017). Configural face processing impacts race disparities in humanization and trust. *Journal of Experimental Social Psychology, 73*, 111-124.

Cassidy, K. D., Boutsen, L., Humphreys, G. W., & Quinn, K. A. (2014). Ingroup categorization affects the structural encoding of other-race faces: Evidence from the N170 event-related potential. *Social Neuroscience, 9*, 235-248.

~~Civile, C., Colvin, E., Siddiqui, H., & Obhi, S. S. (in press). Labelling faces as 'Autistic' reduces the inversion effect. *Autism*, doi: 1362361318807158.~~

Cloutier, J., Mason, M. F., & Macrae, C. N. (2005). The perceptual determinants of person construal: Reopening the social-cognitive toolbox. *Journal of Personality and Social Psychology, 88*, 885-894.

Costello, K., & Hodson, G. (2010). Exploring the roots of dehumanization: The role of animal—human similarity in promoting immigrant humanization. *Group Processes & Intergroup Relations, 13*, 3-22.

Deska, J. C., & Hugenberg, K. (2017). The face-mind link: Why we see minds behind faces, and how others' minds change how we see their face. *Social and Personality Psychology Compass, 11(12)*, e12361.

Fincher, K. M., & Tetlock, P. E. (2016). Perceptual dehumanization of faces is activated by norm violations and facilitates norm enforcement. *Journal of Experimental Psychology: General, 145*, 131-146.

Fincher, K. M., Tetlock, P. E., & Morris, M. W. (2017). Interfacing with faces: Perceptual humanization and dehumanization. *Current Directions in Psychological Science, 26*, 288-293.

Fiske, S. T. (2013). Varieties of (de) humanization: Divided by competition and status. In *Objectification and (de)humanization* (pp. 53-71). Springer, New York, NY.

Freeman, J. B. (2018). Doing psychological science by hand. *Current Directions in Psychological Science. 27,* 315–323.

Freeman, J.B. & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review, 118*, 247-279.

Freeman, J.B. & Ambady, N. (2014). The dynamic interactive model of person construal: Coordinating sensory and social processes. In J. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual Process Theories of the Social Mind*. New York: Guilford Press.

Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General, 137*, 673-797.

Freeman, J.B., Dale, R., & Farmer, T.A. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology, 2*, 59.

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.

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review, 10*, 252-264.

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

Hehman, E., Carpinella, C. M., Johnson, K. L., Leitner, J. B., & Freeman, J. B. (2014). Early processing of gendered facial cues predicts the electoral success of female politicians. *Social Psychological and Personality Science, 5*, 815-824.

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations, 18*, 384-401.

Hugenberg, K., & Corneille, O. (2009). Holistic processing is tuned for in-group faces. *Cognitive Science, 33*, 1173-1181.

Hugenberg, K., & Wilson, J. P. (2013). Faces are central to social cognition. In Carlston, D. (Ed.), *The Oxford Handbook of Social Cognition*. Oxford University Press.

Hugenberg, K., Young, S. G., Rydell, R. J., Almaraz, S. M., Stanko, K., See, P. E., & Wilson, J. P. (2016). The face of humanity: Configural face processing influences ascriptions of humanness. *Social Psychological and Personality Science, 7*, 167-175.

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology, 85*, 616-626.

Judd, C. M., Westfall, J., & Kenny, D. A. (2016). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology, 68*, 601-625.

Kawakami, K., Amodio, D. M., & Hugenberg, K. (2017). Intergroup perception and cognition: An integrative framework for understanding the causes and consequences of social categorization. In *Advances in Experimental Social Psychology* (Vol. 55, pp. 1-80). Academic Press.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. Journal of Statistical Software, 82(13), 1 - 26. doi:http://dx.doi.org/10.18637/jss.v082.i13

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*, 255-260.

McKelvie, S. J. (1995). Emotional expression in upside-down faces: Evidence for configurational and componential processing. *British Journal of Social Psychology, 34*, 325-334.

McKone, E., & Robbins, R. (2011). Are faces special? *Oxford Handbook of Face Perception*, 149-176.

Minear, M. & Park, D.C.(2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers. 36*, 630-633.

Michel, C., Corneille, O. & Rossion, B. (2006). Race categorization modulates holistic face encoding. *Cognitive Science, 31*, 911-924.

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychological Review, 98*, 164-181.

Parkinson, B. (2005). Do facial movements express emotions or communicate motives? *Personality and Social Psychology Review, 9*, 278-311.

Quinn, K. A., & Macrae, C. N. (2005). Categorizing others: the dynamics of person construal. *Journal of Personality and Social Psychology, 88*, 467-479.

R Development Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rhodes, G., Brake, S., Taylor, K., & Tan, S. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology, 80*, 313–331.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition, 25*, 583-592.

Taubert, J., Qureshi, A. A., & Parr, L. A. (2012). The composite face effect in chimpanzees (Pan troglodytes) and rhesus monkeys (Macaca mulatta). *Journal of Comparative Psychology, 126*, 339-346.

Wiese H., Stahl J., Schweinberger, S.R. (2009) Configural processing of other-race faces is delayed but not decreased. *Biological Psychology, 81*, 103–109

Wilson, J. P., Young, S. G., Rule, N. O., & Hugenberg, K. (2018). Configural processing and social judgments: Face inversion particularly disrupts inferences of human-relevant traits. *Journal of Experimental Social Psychology, 74*, 1-7.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79*, 471-491.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*, 141-145.

Young S.G., Bernstein, M.J., & Hugenberg, K. (2010). When do Own-Group Biases in face recognition occur? Encoding versus post-encoding. *Social Cognition, 28*, 240-250.

Young, S.G., Hugenberg, K., Bernstein, M.J., & Sacco, D.F. (2009). Intergroup salience decreases recognition for same-race faces. *Journal of Experimental Social Psychology, 45*, 1123-1126.

Young, S.G., Goldberg, M., Rydell, R.J., & Hugenberg, K. (2019). Trait anthropomorphism predicts ascribing human traits to upright but not inverted chimpanzee faces. *Social Cognition, 37*, 105-121.

Young, S. G., & Hugenberg, K. (2010). Mere social categorization modulates identification of facial expressions of emotion. *Journal of Personality and Social Psychology, 99*, 964-977.

Zhao, M., Hayward, W. G., & Bülthoff, I. (2014). Holistic processing, contact, and the other-race effect in face recognition. *Vision Research, 105*, 61-69.

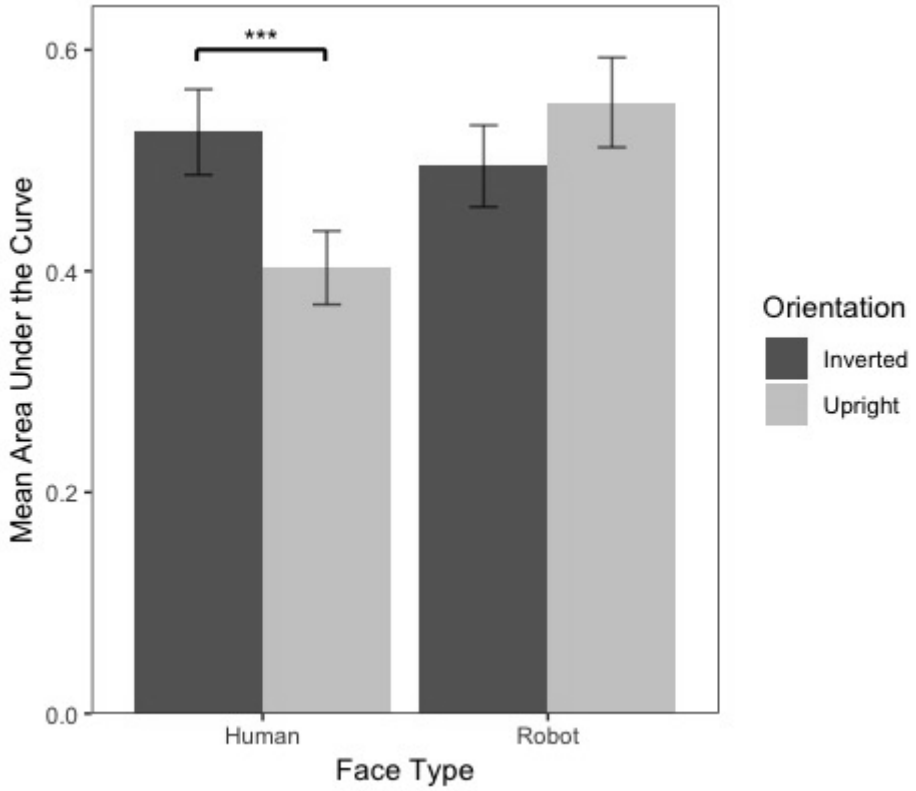

Figure 1. Mean AUC values for inverted versus upright human and robot faces in Experiment 1. Error bars represent 95% CI for mean AUC values. *** indicated $p < .001$

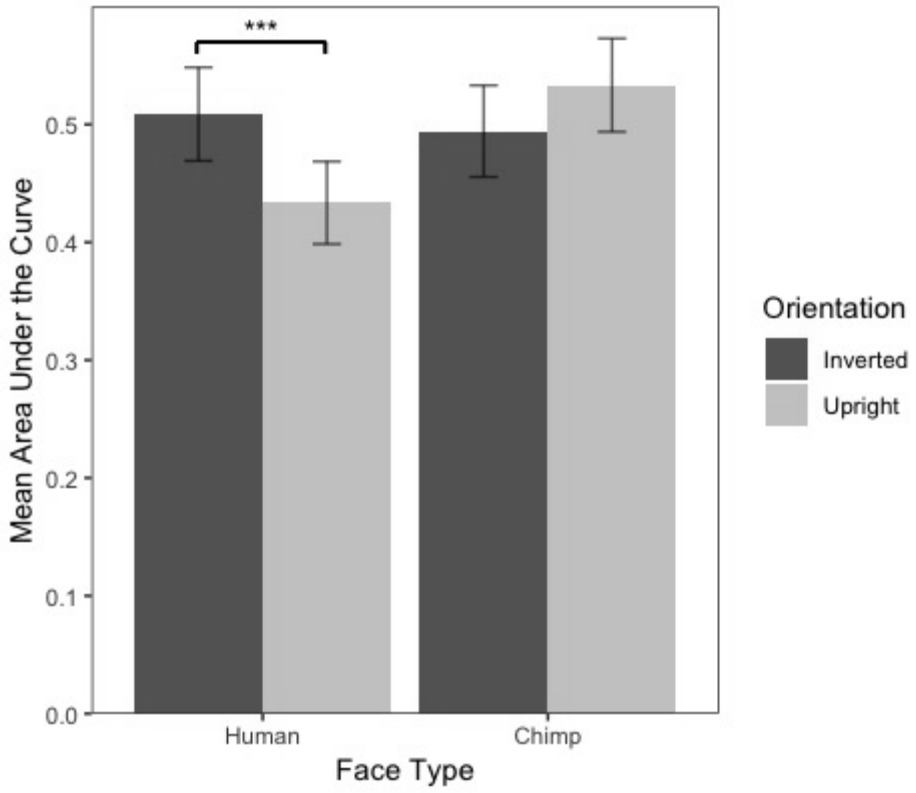

Figure 2. Mean AUC values for inverted versus upright human and chimp faces in Experiment 2. Error bars represent 95% CI for mean AUC values. *** indicated $p < .001$

Supplementary Material

Below are the results from supplementary analyses that examined common dependent variables produced by MouseTracker, including maximum deviation (MD), number of flips across the x-axis (x-flips), and log-transformed reaction times (RT). These analyses are based on the same final samples collected in the originally reported results and all exclusion criteria are identical to those reported in the main text.

Study 1

Maximum Deviation

We first examined each participant's MD value, defined as the most extreme point their mouse trajectory takes from the ideal straight-line trajectory. Participants' MD scores were analyzed in a cross-classified linear mixed effects model specifying random factors for participants and stimuli and regressed Orientation (upright = .5, inverted = -.5) and Face Type (human = .5, robot = -.5) and their interaction on MD. Model fit was determined in a similar fashion as described in the main section, where a model specifying random intercepts for participants and stimuli with a random slope for Face Type provided the best fit to the data, $\chi^2(3) = 282.31$, $p < .001$. In this model, the random intercept for participants was significant ($SD = .15$, 95% CI [.13, .19]), as was the random slope for Face Type ($SD = .19$, 95% CI [.16, .24]) and the random intercept for stimuli ($SD = .02$, 95% CI [.01, .04]). Results revealed a marginally significant main effect for Orientation, $b = -.02$, $SE = .01$, $t(56.13) = -1.96$, $p = .055$, 95% CI [-.05, -.0001], and a nonsignificant main effect for Face Type ($p = .36$). There was a significant Orientation x Face Type interaction, $b = -.09$, $SE = .02$, $t(56.13) = -3.79$, $p < .001$, 95% CI [-.13, -.04] (see Figure S1). To examine this interaction, we split the data based on Orientation and then built additional linear mixed effects models regressing MD on Face Type. Model comparison analyses revealed that, compared to only a model with a random intercept with

participants, a model specifying a random intercept for participants with a random slope for Face Type provided the best fit for the data, $\chi^2(2) = 149.5$, $p < .001$. Significance tests revealed that the random intercept for participants was significant, *SD* = .16, 95% CI [.13, .19], as was the random slope, *SD* = .22, 95% CI [.17, .27]. Results revealed that for upright targets, participants tended to have a significantly lower MD value for human versus robot faces, *b* = -.07, *SE* = .03, *t*(59.75) = -2.27, *p* = .03, 95% CI [-.13, -.009]. The model for inverted targets revealed no significant differences for human versus robot faces.

X-flips

X-flips, defined as the number of reversals participants have along the x-axis, is another common MouseTracker measure often used as a measure of preference reversal. We determined the appropriate random effects structure using the same procedure as described in the MD analyses. Of all models, the only model that was not overfit to the data was the model including a random intercept for participants. The random effects structure indicated that participants did account for a significant proportion of variance, *SD* = 1.15, 95% CI [.96, 1.39]. This model revealed nonsignificant main effects for Orientation and Face Type (*p*s < .41). There was a significant Orientation x Face Type interaction, *b* = -.28, *SE* = .13, *t*(7067.14) = -2.21, *p* = .03 , 95% CI [-.53, -.03] (see Figure S2). Examining this interaction by Orientation revealed only a marginal effect of Face Type for upright targets, *b* = -.19, *SE* = .11, *t*(60.53) = -1.80, *p* = .08, 95% CI [-.41, .02]. No effect was found for inverted targets.

Reaction Time

Our next analyses considered participants' log-transformed RT values as a metric of how quickly they arrived at their final categorization. Participants' log-transformed RTs were examined using the same cross-classified linear mixed effects model, where model comparison

tests revealed that a model specifying random intercepts for participants and stimuli and a random slope for Face Type provided the best fit compared to models with simpler random effects specifications, $\chi^2(2) = 44.81$, $p < .001$. The random intercept for participants was significant, $SD = .17$, 95% CI [.15, .21], as was the random slope for Face Type, $SD = .04$, 95% CI [.03, .06], and the random intercept for stimuli, $SD = .02$, 95% CI [.02, .03]. Here, we found only a marginal main effect of Orientation, $b = -.01$, $SE = .01$, $t(56.25) = -1.90$, $p = .06$, 95% CI [-.03, .0003].

Study 2

Maximum Deviation

As with study 1, we first examined participants' MD scores by regressing these values on Orientation (upright = .5, inverted = -.5) and Face Type (human = .5, chimp = -.5) and their interaction using a linear mixed effects model. Model comparison tests revealed that a model specifying random intercepts for participants and stimuli and a random slope for Face Type provided the best fit over models with simpler random effects structures, $\chi^2(3) = 141.42$, $p < .001$. Significance tests for the random effects structure revealed that the random intercept for participants was significant, $SD = .15$, 95% CI [.13, .18], as was the random slope for Face Type, $SD = .16$, 95% CI [.13, .20]. The random intercept for stimuli was not significant, $SD = .01$, 95% CI [.00, .03], though because the model including this was not overfit, it was left in. Results from the model revealed nonsignificant main effects for both Orientation and Face Type ($p = .14$, $p = .62$, respectively). The Orientation x Face Type interaction was significant, $b = -.07$, $SE = .02$, $t(73.30) = -3.40$, $p = .001$, 95% CI [-.11, -.03] (see Figure S4). Breaking this interaction down by Orientation revealed that a model specifying a random intercept for participants with a random slope for Face Type provided the best fit, $\chi^2(2) = 55.76$, $p < .001$. Results revealed that for upright targets, human faces had significantly lower MD scores compared to chimp faces, $b = -$

.04, *SE* = .02, *t*(84.95) = -2.00, *p* = .049, 95% CI [-.09, -.001]. No effect was found for chimp faces.

X-flips

We next examined participants' x-flips using the same model specified above. Model fit tests revealed that a model specifying random intercepts for participants and stimuli with a random slope for Face Type significantly improved fit over simpler models, $\chi^2(0) = 3.28$, $p < .001$. In this model, the random intercept for participants was significant, *SD* = 1.24, 95% CI [1.05, 1.45], as was the random slope for Face Type, *SD* = .32, 95% CI [.002, .53]. The random intercept for stimuli was not significant, *SD* = .19, 95% CI [.00, .28], but because the model was not overfit it was left in. Here we found only a marginally significant main effect of Face Type, *b* = -.16, *SE* = .09, *t*(59.48) = -1.83, *p* = .07, 95% CI [-.33, .01], indicating that participants had significantly fewer x-flips for human targets than for chimp targets (see Figure S5). The model revealed a nonsignificant main effect for Orientation and a nonsignificant interaction (*p* = .40, *p* = .13).

Reaction Time

We constructed the same model used in the previous analyses on log-transformed RTs. Our analyses revealed that a model specifying random intercepts for participants and stimuli with a random slope for Face Type significantly improved fit over simpler models, $\chi^2(3) = 57.84$, $p < .001$. In this model, the random intercept for participants was significant, *SD* = .20, 95% CI [.17, .23], as was the random slope for Face Type, *SD* = .04, 95% CI [.03, .06], and the random intercept for stimuli: *SD* = .02, 95% CI [.01, .03]. Here, only the main effect of Orientation was significant, *b* = -.03, *SE* = .01, *t*(73.60) = -4.14, *p* < .001, 95% CI [-.04, -.02], such that participants were faster at categorizing upright versus inverted targets (see Figure S6). Neither the main effect of Face Type nor the interaction were significant (*p* = .65, *p* = .10, respectively).

Supplementary Appendix

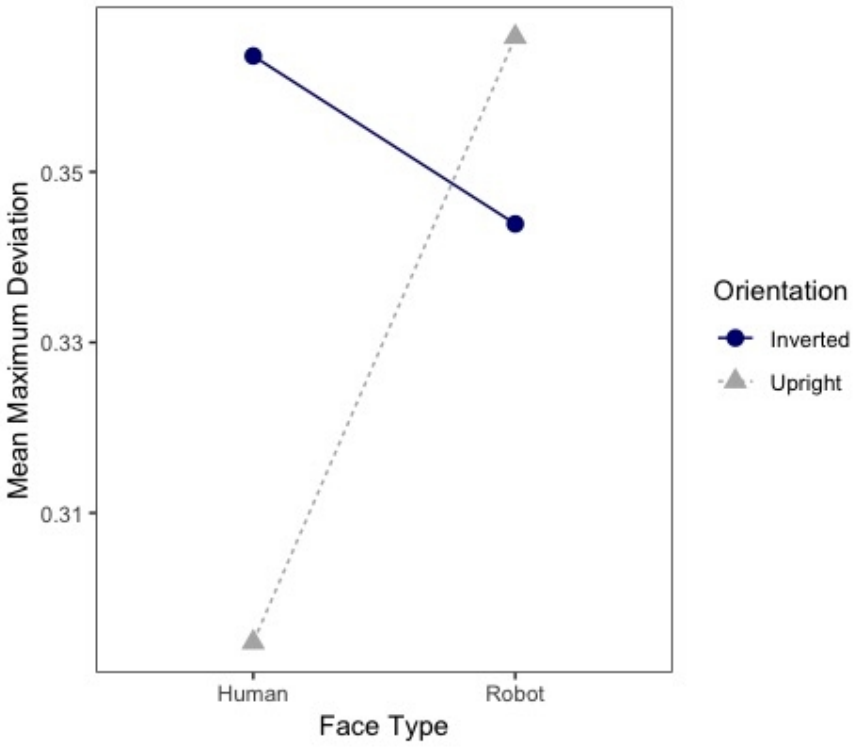

Figure S1. Interaction plot for maximum deviation (MD) analysis for Study 1.

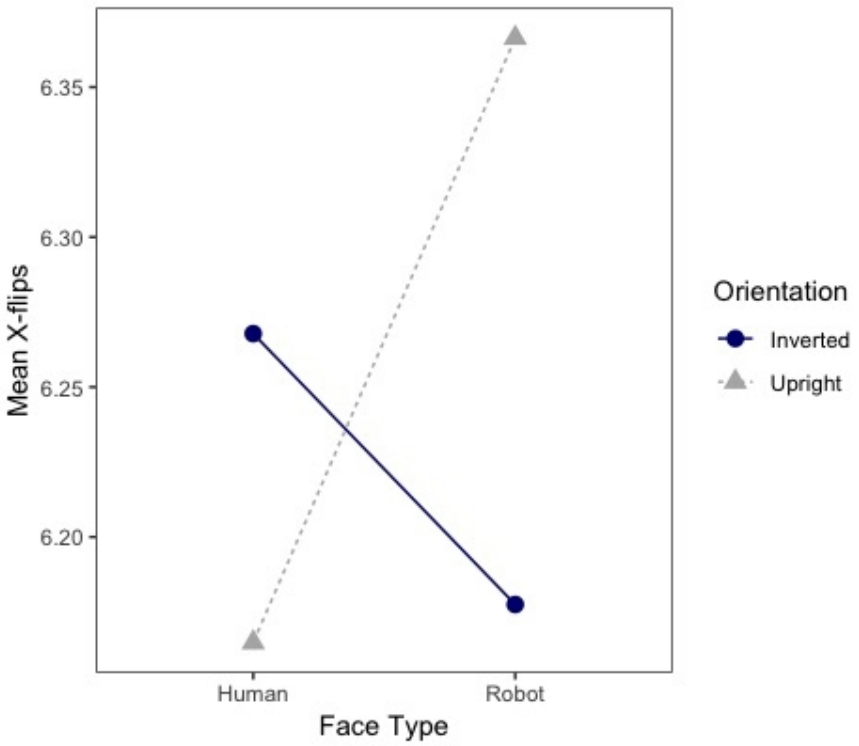

Figure S2. Interaction plot for the x-flips analysis for Study 1.

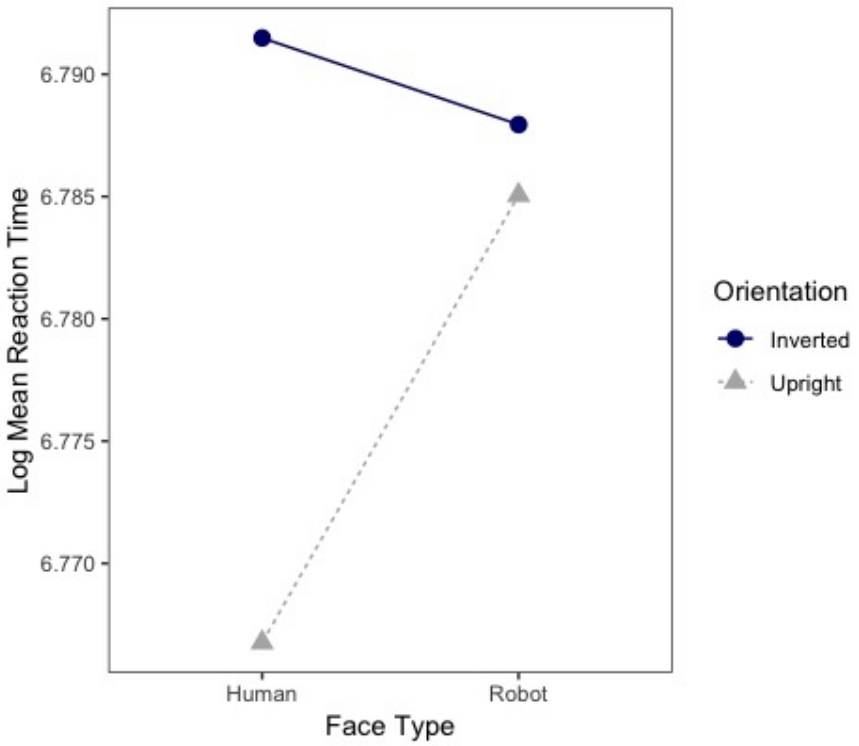

Figure S3. Interaction plot for the log-transformed reaction time (RT) analysis for Study 1.

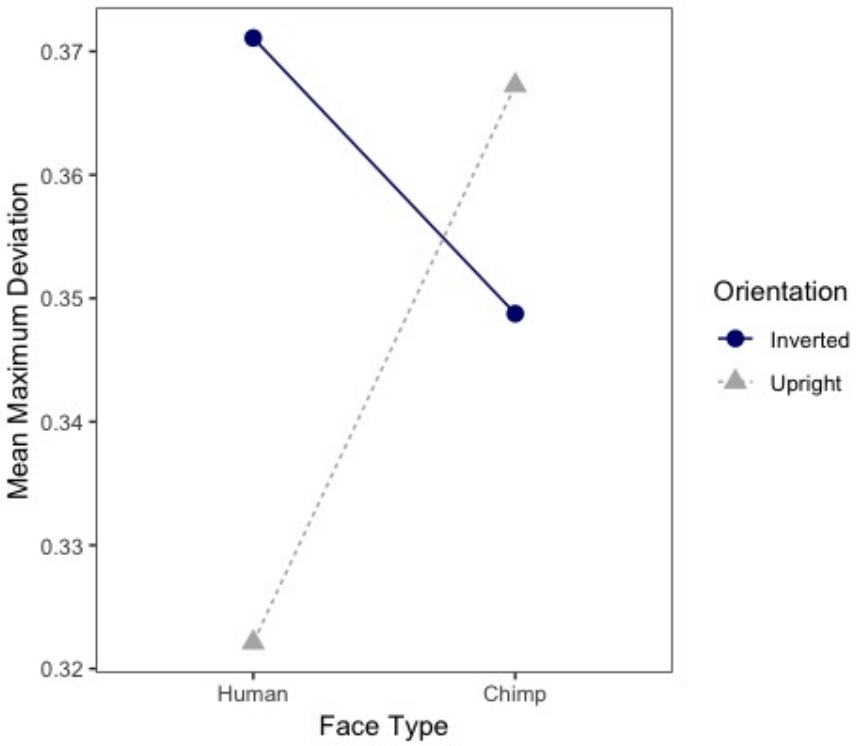

Figure S4. Interaction plot for maximum deviation (MD) analysis for Study 2.

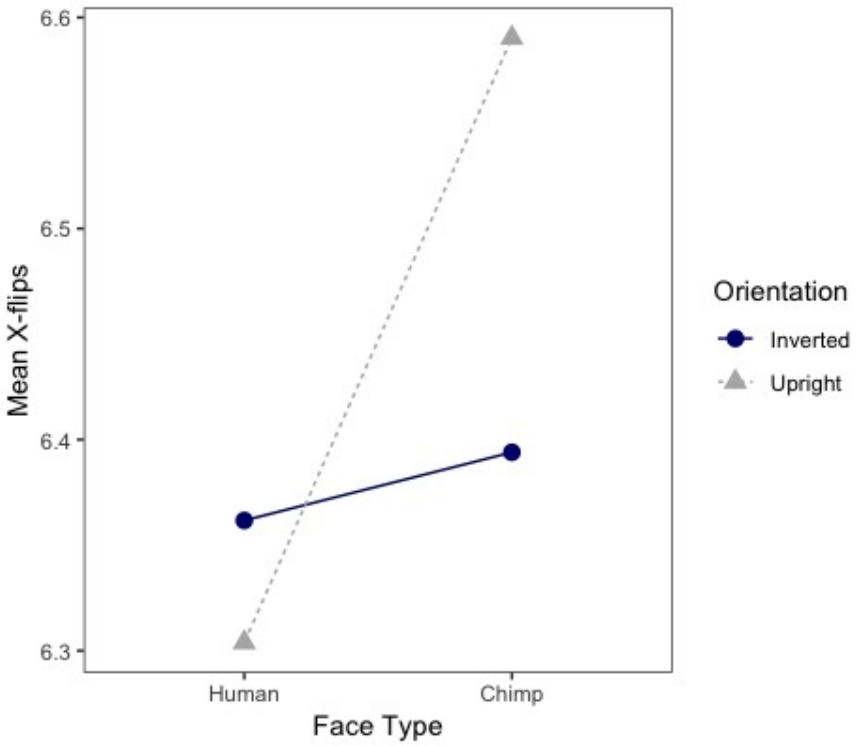

Figure S5. Interaction plot for x-flips analysis for Study 2.

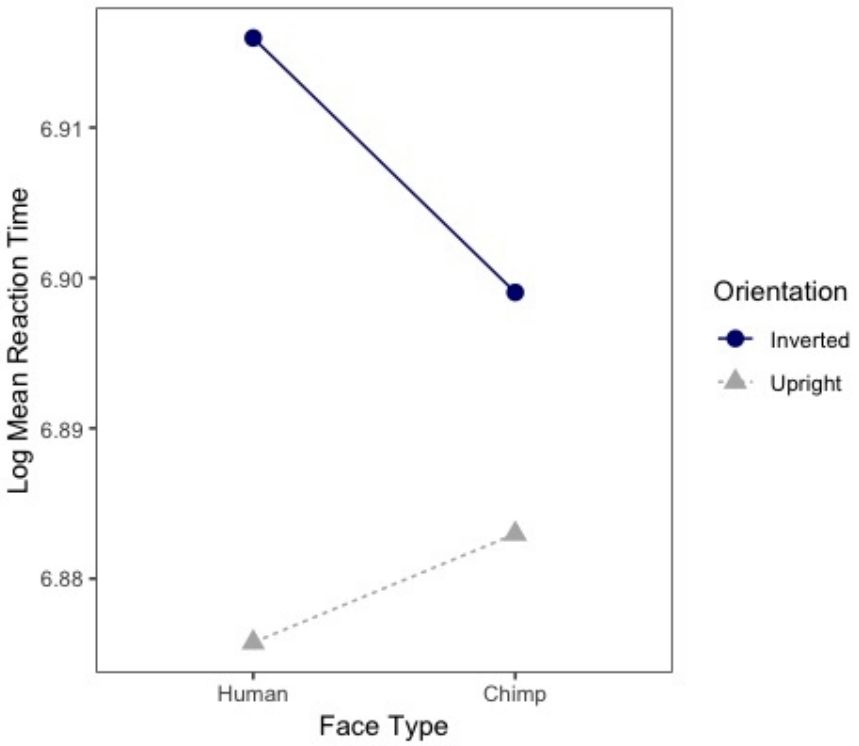

Figure S6. Interaction plot for the log-transformed reaction time (RT) analysis for Study 2.