The hidden cost of humanization: Individuation reduces prosocial behavior toward ingroup members

Victoria K. Lee[1]

Rachel Kranton[2]

Scott A. Huettel[1,3]

[1]Department of Psychology and Neuroscience

[2]Department of Economics

[3]Center for Cognitive Neuroscience

Duke University

Abstract

Humanization – the recognition of another individual's internal mental life – has been proposed as an effective method for increasing prosocial behavior. Specifically, the engagement of social cognitive processes causes previously dehumanized outgroup members to be considered as individuals, reducing the typical ingroup/outgroup bias. The consequences of humanization for ingroup members, however, are less clear: Does it reinforce positive ingroup identities and lead to increased pro-social behavior? Or, does it compete with those identities, and thus counterintuitively reduce pro-social behavior? We conducted two studies in which participants performed incentive-compatible dictator games to allocate money to real people based on knowledge of group membership. We manipulated humanization by providing one line of additional information that either reinforced the group membership or humanized the recipient. We found robust support for a competitive model of humanization in which group and individual identities compete; specifically, humanizing information increased pro-social behavior toward outgroup members but decreased pro-social behavior toward ingroup members. Furthermore, we demonstrated that this effect results from a decreased reliance on group membership labels and increased reliance on similarity ratings, and not from the induction of new group identities. Our results were replicated in both political and experimentally induced groups and in two separate experiments – and were evident within individual subjects. We conclude that humanization can carry a hidden cost for ingroup members by disrupting group identities that would otherwise make them the target of altruistic actions.

Keywords: prosocial behavior, group bias, humanization,

Significance Statement

Humanization is advantageous to prosocial behavior. Thinking about the minds of others encourages altruistic behavior (e.g. helping, charitable giving) and reduces intergroup biases. Here, we highlight an important boundary condition to these benefits – demonstrating that humanizing ingroup members decreases, rather than increases, prosocial giving in an incentive-compatible dictator game. Using multi-level modeling, we show that this reduction results from a change in the weighting placed on group identity versus personal similarity, with a clear and large-magnitude double dissociation observed. These results have important implications for social policy by suggesting that interventions should humanize outgroup members, but not ingroup members, to maximize prosocial behaviors like altruistic giving.

The hidden cost of humanization: Individuation reduces prosocial behavior toward ingroup members

\body

Prosocial behavior is a cornerstone for societal well-being (1) and understanding its determinants has been a cardinal goal for social and behavior science (2,3). Through the use of behavioral studies, neuroimaging, and meta-analyses, it is becoming increasingly clear that increased social cognitive processing (e.g. mentalizing and perspective taking) is associated with greater prosocial behavior (e.g. charitable giving, cooperation, helping) in both children and adults (4-8). Importantly, these social cognitive processes contribute to the perception of others as humans with active mental states (9), suggesting this process of *humanization* may provide an ideal strategy for increasing prosocial behavior (10-12).

Humanization is particularly relevant for intergroup relations, as group identity is often associated with differences in humanized perception and prosocial behavior. Outgroup members often suffer the consequences of dehumanized perception (e.g., a failure to spontaneously infer mental states), characterized by reduced activity in brain regions involved in social cognition when perceiving outgroup members (13), which results in the attribution of fewer social emotions to outgroup members (14, 15) and a reduced willingness to help and increased willingness to harm members of the outgroup (16, 17). In contrast, the humanized perception of ingroup members (18) is characterized by the engagement of brain regions associated with social cognition (19-23), resulting in the greater liking of ingroup members (24), a higher likelihood of helping ingroup members (25), and ingroup favoritism during resource allocation (26). Moreover, manipulations of social cognition can reduce intergroup bias across a variety of contexts, including humanizing traditionally dehumanized outgroup members (27), reduced stereotype activation (28), more equitable social norm enforcement (29, 30), and attenuated parochial empathy (31). Collectively, these results suggest that intergroup biases in prosocial behavior may occur because the benefits of humanization (see 32, 33) are naturally afforded to ingroup members but denied to outgroup members.

That humanization reduces intergroup biases has become a well-accepted phenomenon. Yet, the individual effects that contribute to this reduced bias have yet to be

explored in detail. That is, reduced intergroup bias in prosocial behavior could be explained by (at least) three distinct humanization models, each with different implications for its underlying mechanisms (Figure 1). First, given evidence that engagement of social cognitive processes is associated with greater prosocial behavior (4-8), a *generalized humanization* model would predict that humanization increases prosocial behavior toward both ingroup and outgroup members. Second, humanization could only affect outgroup members since research suggests ingroup members are already humanized (19-23). Such an *outgroup humanization* model predicts that humanization would lead to increased prosociality toward outgroup members but have no effect toward ingroup members. Third, and perhaps most counterintuitively, humanization could focus attention away from categorical group membership and its associated benefits (34) and toward a more individual, humanized identity (31, 35). Such a *competitive humanization* model would predict an increase in prosociality toward outgroup members but a decrease in prosociality toward ingroup members. This result, if observed, would demonstrate an important boundary condition to the benefits of humanization – with corresponding implications for social policy.

To test the differential predictions of these models, we measured prosocial behavior using sets of incentive-compatible dictator games in which participants allocated points to other people. After completing a non-group control condition, participants and their giving targets were divided into groups based on preferences related to real-world issues (political groups) or arbitrary art preferences (minimal groups). We then provided additional information that either reinforced the group membership (categorical condition) or encouraged thinking about the target's goals and preferences (humanized condition; see Figure 2). We used a fully within-subjects design that not only provided direct comparisons between different group and humanization conditions, but also enabled replication across two distinct group manipulations. And, we tested our key results across two experiments with independent groups of subjects, while also collecting additional data about how the humanizing statements reinforced individual or group identities. This approach allowed robust and multiply-replicated tests of the different humanization models, while ruling out alternative explanations for our results.

Across two studies, we found strong support for the competitive humanization model: humanizing information increases prosocial behavior toward outgroup members and decreases prosocial behavior toward ingroup members. Through analysis of item-specific effects, we observed a clear double dissociation. When humanizing information was absent, prosocial giving was influenced by group membership but not by perceived similarity between the giver and target. But, when humanizing information was present, group membership has minimal effects while individual similarity dominates giving. Additional control analyses rejected alternative explanations for the observed pattern of results, including social distance and the creation of a second group identity. Together these results support the notion that group and individual identities compete to determine prosocial behavior – revealing that humanization does not simply benefit pro-social behavior, but also can carry a hidden cost.

Study 1: Results and Discussion

To examine the effects of humanization on prosocial behavior separately for ingroup and outgroup members, participants allocated points in dictator games to group members under categorical or humanized perception. For each trial, we calculated the percentage of points allocated to the other person. Averages for each condition were entered into a 2 (group: own/other) x 2 (perception: categorical/humanized) repeated-measures ANOVA. In both the political groups and minimal groups manipulations, we observed a significant group x perception interaction, (political groups: $F(1, 75) = 35.625$, $p < .001$, partial $\eta^2=.322$; minimal groups: $F(1, 75) = 25.719$, $p < .001$, partial $\eta^2=.255$). T-tests were conducted following the significant interactions and are reported below.

Political groups. Consistent with literature demonstrating that people feel both an "ingroup love" and "outgroup hate" toward members of different groups (34), participants in the political group context gave more points to categorical ingroup members and fewer points to categorical outgroup members, compared to nongroup controls (see Table 1; ingroup: $t(75) = 3.351$, $p =.001$; outgroup: $t(75) = -5.12$, $p <.001$). Thus, under the categorical perception of group members, we replicated the classic

intergroup bias effect in which participants give more points to categorical ingroup members than categorical outgroup members, $t(75) = -6.892$, $p < .001$.

Importantly, these group biases were attenuated by a simple manipulation: providing one sentence of humanizing/individuating information about the target person. Participants still give more to humanized ingroup members than humanized outgroup members, $t(75) = -4.158$ $p < .001$, but that difference is about one quarter as large as in the categorical condition. This manipulation not only increased allocations to humanized outgroup members, $t(75) = -4.106$, $p < .001$, but actually decreased allocations to humanized ingroup members $t(75) = 4.528$, $p < .001$; see Figure 3a.

Minimal groups. When groups were formed based on arbitrary preferences (i.e., our minimal group manipulation), we observed similar results to the political group manipulation. We again observed a group bias where participants allocated more to categorical ingroup members than categorical outgroup members, $t(75) = -4.457$, $p < .001$, and this bias was eliminated under humanized perception, with participants allocating similar amounts to humanized ingroup members and humanized outgroup members, $t(75) = -0.618$, $p = .518$; see Table 1. This reduction of the group bias resulted from a decrease in allocations to ingroup members when changing perception from categorical to humanized, $t(75) = 4.564$, $p < .001$. However, unlike the political groups, we did not observe a significant increase when changing perception of outgroup members, $t(75) = -1.256$, $p = .213$. Additionally**,** here categorization as an outgroup member resulted in less prosocial behavior as compared to control, $t(75) = 2.559$, $p = .013$, whereas individuals categorized as members of an ingroup were treated similarly to the nongroup control ($p > 0.05$).

Collectively, these results supported the competitive humanization model in which group and individual identities compete to determine prosocial behavior. Despite growing evidence that suggests social cognitive processing is associated with increased prosocial behavior, we observed a decrease in prosocial behavior for humanized ingroup members in both the political and minimal groups manipulations. In study 2, we tested whether these results replicate while also ruling out alternative explanations, such as the effects of social distance (6, 36-37) and a second group identity. While repeating our basic paradigm, we also collected new data (similarity ratings) that provide a novel test

for the competitive humanization model. Specifically, if group and individual identities compete under humanized perception, reliance on group membership labels should decrease while reliance on individual similarity ratings should increase when going from categorical to humanized perception.

Study 2: **Results and Discussion**

Study 2 fully replicated the effects observed in study 1—we observed a significant group x perception interaction for both political and minimal group manipulations (political groups: $F(1, 75) = 36.67$, $p < .001$, partial $\eta^2 = .328$; minimal groups: $F(1, 75) = 20.144$, $p < .001$, partial $\eta^2 = .212$), with t-tests for the simple effects reported below.

Political groups. Like study 1, we observed a categorical group bias in which participants gave more to ingroup members than outgroup members, $t(75) = -7.521$, $p < .001$. This bias remained intact for the humanized conditions, but was largely mitigated $t(75) = -4.633$, $p < .001$; see Figure 3b. Replicating the effects from study 1, this attenuation resulted from an increase in allocations for humanized outgroup members, $t(75) = -6.117$, $p < .001$, and a decrease in allocations for humanized ingroup members, $t(75) = 3.958$, $p < .001$. Compared to the control condition, only outgroup members received significantly less points in the categorical, $t(75)=6.838$, $p < .001$, and humanized conditions, $t(75)=3.821$, $p < .001$, although the categorical ingroup condition was marginally different from control and in the same direction as observed in study 1, $t(75)=- 1.685$, $p= .096$.

To examine the robustness of this effect across individual participants, we plotted all participants' ingroup bias scores (i.e., ingroup minus outgroup allocations) for both the categorical and humanized conditions (see Figure 4, which includes all participants in both studies). The effects were strikingly consistent across participants, such that nearly all participants who showed an ingroup bias exhibited strong attenuation or elimination of that bias when provided humanizing information.

Minimal groups. We again replicated the basic pattern of results in the minimal group condition. Under categorical perception, participants were more prosocial toward ingroup members than outgroup members, $t(75) = -6.463$, $p < .001$, and the group difference remained intact for the humanized conditions, with participants giving more to

humanized ingroup members than humanized outgroup members, $t(75) = -2.583$, $p = .012$. Additionally, changing perception from categorical to humanized for ingroup members decreased the number of points allocated, $t(75) = 2.147$, $p = 0.035$, but we did not observe a significant increase for humanized outgroup members, $t(75)=-.891$, $p = .376$ (see Table 1). Similar to the political groups, only outgroup members received significantly less points in the categorical, $t(75)=2.392$, $p = .019$, and humanized conditions, $t(75)=2.537$, $p = .013$, than the control condition. The ingroup members were not significantly different for the control condition under categorical or humanized perception ($p > .05$).

Similarity. Because people may give more to individuals similar to themselves than to dissimilar individuals, we wanted to evaluate the extent to which similarity of the humanizing statements drove our results. That is, in the categorical conditions similarity judgments can only based on group membership, since all of the information reinforced that identity. But in the humanized conditions, participants may make a similarity judgment based on group identity, and then adjust those judgments based on the degree of similarity of the humanizing information. Therefore, in study 2 we collected similarity ratings for each of the statements and used multilevel modeling to examine the extent to which group identity and similarity ratings differentially guide allocation decisions under categorical and humanized perception (see Methods).

Using three nested models (baseline, group, similarity), we first examined whether similarity ratings explain additional variance above and beyond group identity. Consistent with the group effect reported above, group identity significantly improved model fit compared to the baseline model for both categorical and humanized perception (categorical: $X^2(1) = 1019.07$ $p < .001$; humanized: $X^2(1) = 87.05$, $p < .001$). Adding similarity ratings further improved model fit in both the categorical and humanized conditions (categorical: $X^2(1) = 65.12$, $p < .001$; humanized: $X^2(1) = 405.89$, $p < .001$), suggesting that the similarity of the statements to the participant plays a role in allocation decisions (see Table 2). Additionally, by calculating the change in residual variance ($\Delta\sigma^2$) at each step, we were able to see how reliance on group identity and similarity changes under humanized perception. Under categorical perception, group identity accounts for approximately 24% of the residual variance (see Table 2). However, similarity ratings for

the categorical statements only explained an additional 1.3% of the residual variance. Under humanized perception, we found the opposite pattern—group identity only accounts for approximately 1.9% of the residual variance, while similarity ratings explained an additional 8.5% of the residual variance.

Together, these results suggest a clear double dissociation. When participants thought about another person as a member of a group, their dictator game allocations were based much more on group identity (24% of residual variance) than on similarity to oneself (1.3% of residual variance). However, when the target was humanized, then similarity had a much larger effect than group identity (8.5% versus 1.9% of the residual variance; see Table 2). Importantly, the reduced reliance on group identity when moving from categorical ($\Delta\sigma^2$ = 24%) to humanized ($\Delta\sigma^2$ = 1.9%) also supported the conclusion that group and individual identities compete under humanized perception, such that humanizing information dramatically diminishes the effect of group identity on pro-social decisions.

Reinforcing Individual vs. Group Identities. As an additional control analysis, we wanted to ensure that the humanizing statements were in fact individuating the recipient and not just creating a second group identity. That is, rather than humanizing the recipient, the statements could evoke a second ingroup or outgroup identity based on the similarity of the humanizing statement to the participant (e.g., the statement "if I could own any animal in the world it would be a dog" could be encoded as "we are both 'dog people'"). To evaluate this possibility, an independent participant sample provided "humanization rating" (range 0-100), in which statements that evoke group identities were given lower numbers and statements that evoke individual identities were given higher numbers.

On average, the statements were judged as evoking a more individual identity (M = 67.5, SD = 7.73). This provides a manipulation check to ensure that the humanizing statements were in fact individuating the recipient. However, the statements varied in the extent to which they evoked a group versus individual identity, ranging from 45.68 (most group-like: "When I was little I wanted to be a football player when I grew up") to 84.61 (most individuating: "The most interesting thing about me is I pull out my own rotten teeth"). Thus we entered this "groupiness" rating into a multilevel model that included

the original group membership label (own/other), the participant's similarity rating of the humanizing statement, the statement's groupiness rating (from Mturk), and a similarity x groupiness interaction to see whether these additional variables would eliminate our group effects. Our analysis revealed that the effect of the original group membership labels (own/other) was still significant even when controlling for the similarity and groupiness of the statement (political groups: $t(4454.02)=9.09$, $p < 0.001$; minimal groups: $t(4481.001) = 3.575$, $p < 0.001$), suggesting that these additional variables do not explain away the group effect.

Additionally, if the humanizing statements were creating a second group identity, we might expect to find a significant interaction such that there is a stronger effect of similarity for the more group-like statements (i.e. a second ingroup/outgroup effect) compared to the more individuating statements. However, we found no evidence to suggest that this is the case and in fact found evidence for the opposite. For the political groups, the similarity x groupiness interaction was not significant ($t(4455.38) = 1.542$, $p = .123$). For the minimal groups, we did find a significant interaction ($t(4483.67) = 2.732$, $p = .006$), however, upon probing the interaction our results revealed that it was the more individuating statements and not the more group-like statements that had a stronger effect of similarity (individuating statements: $b_{similarity} = .15$, $t(4486.18) = 12.756$, $p < 0.001$; group-like statements: $b_{similarity} = .09$, $t(4484.23) = 8.746$, $p < 0.001$). Together these results support the idea that the humanizing statements were in fact individuating the recipient and not creating a second group identity.

General Discussion

Our experiments revealed two consistent effects in pro-social decision-making. First, consistent with previous research, when only information about group membership was available (i.e. under categorical perception), participants were more generous to members of their ingroup than to members of an outgroup. This result replicated across two independent experiments and across political and minimal groups – with stronger effects for the real-world category of political groups than for the laboratory-induced minimal-group (34). Second, and perhaps more strikingly, we found that humanizing information has differential effects on pro-social behavior: it increases pro-social giving to outgroup members, but decreases pro-social giving to ingroup members. Interestingly,

this decrease for ingroup members contradicts the literature suggesting social cognition is associated with increased prosocial behavior. However, this effect was also replicated across two studies while also ruling out alternative explanations. Importantly, then, this suggests there is an important boundary condition to the benefits of humanization.

Together, these results support the competitive humanization model where perceived individual identities compete with perceived group identities to determine pro-social behavior. First, if group and individual identities had been additive rather than competitive, we would expect to see an increase, not the observed decrease, in prosocial behavior for humanized ingroup members. Second, and perhaps more convincingly, the double dissociation in our similarity analysis from study 2 highlights that under humanized perception, participants simultaneously become less reliant on group membership information and more reliant on the similarity ratings related to the recipient's individual identity. These results are supported by recent work showing that providing descriptive narratives reduces parochial empathy by interfering with the encoding of targets' group membership (31). Here we show a similar effect within the context of prosocial behavior, while also highlighting that this reduction occurs from a decrease in prosocial behavior for ingroup members and an increase in prosocial behavior for outgroup members.

Humanization, therefore, not only provides benefits but also carries costs. By competing with group identities it can weaken associated intergroup biases (30), leading people to behave more pro-socially toward humanized outgroup members than toward generic outgroup members. The cost of humanization, however, appears in behavior toward ingroup members; specifically, thinking about an ingroup member as an individual *reduces* prosocial behavior, especially if the humanizing information makes them seem dissimilar to oneself. Under some circumstances, then, dehumanization may be preferred over humanization, leading to better decisions made in a medical (39) or economic context (40). Our results suggest that dehumanized perception of ingroup members, rather than humanized perception, increases prosocial behavior.

Finally, our results contribute to the growing literature on how perception shapes charitable giving – and make strong recommendations for interventions to promote pro-social behavior. Consider charitable giving following a major world disaster (e.g., the

2004 tsunami that devastated much of southeastern Asia). The victims of such events may naturally be seen as dissimilar others: they live in a distant land, speak a different language, engage in different cultural practices, and have just experienced a profound tragedy. Therefore, to promote charitable giving, relief agencies typically try to humanize such large-disasters by describing identifiable, individual victims (i.e., humanizing these outgroup members, 41, 42), consistent with the benefit of humanization shown in the current study. Yet, another approach could be even more effective: emphasizing a common ingroup identity. A fisherman who lost his boat during the tsunami could be described in prosaic terms: Right before the tsunami hit, he had just argued with his daughter about her cell phone use – and had sat down to have breakfast while reading the newspaper. These sorts of descriptors emphasize the commonality of human experience (43) without priming dissimilar group identities. Managing such competition between group identities and individual identities will be critical for encouraging pro-social behavior.

Methods and Materials

Study 1 Methods

Stimulus creation: Participants and Procedure. To eliminate deception and ensure that all social decisions targeted a real person, we conducted a preliminary survey that identified recipients for a two-player Dictator game. One hundred participants were recruited from Amazon's Mechanical Turk (MTurk) to participate in a five to ten minute online survey about preferences. The average age was 32.3 years old (SD = 9.93) and 45% were female. Participants were paid $1 for completing the survey and were eligible to earn additional money based on the actions of participants in the subsequent dictator game experiments.

Following informed consent, participants were asked to indicate their political, art, and personal preferences in an online survey. The political preference questions – which were used to construct meaningful ingroup/outgroup divisions – consisted of five political issues relevant at the time the study was conducted: gay marriage, abortion, gun control, death penalty, and government assistance. The art preference questions were used to construct minimal ingroup/outgroup divisions; they consisted of five pairs of abstract paintings (Klee and Kandinsky) and five pairs of lines of poetry, all matched *a*

priori for content and style. To ensure that the minimal groups were uncorrelated with political groups, we excluded from further use three questions for which one political group tended to prefer one art item over the other. The personal preference questions consisted of seven free-response questions like those used as "icebreakers" at social gatherings (e.g., "If you could own any animal in the world, what would it be?").

At the end of the survey, participants provided demographic information and were given the option to release their anonymous data to be used in future studies. Releasing the data allowed for the opportunity to win bonus money based on future participants' decisions; all participants agreed to release their data and thus serve as a potential target for altruistic giving.

Dictator Game: Participants. Eighty-one potential participants were recruited from Duke University and the surrounding community area, in order to meet an *a priori* target of 75 participants. Data from five individuals were excluded from analyses due to failure to follow instructions (e.g. transferring more than 100% of points) or expressing confusion regarding the rules of the game during the debriefing session (e.g. the independence of one-shot dictator game trials). The final dataset consisted of 76 participants. The average age of the sample was 32.5 years old (SD = 13.7 years) and 64% were female. Participants were paid $12 per hour plus bonus money from a few randomly selected trials (up to an additional $4.59). Study sessions were self-paced and lasted for 1 to 1.5 hours.

Dictator game: Procedure. All procedures and methods were conducted in accordance with guidelines approved by the Institutional Review Board at Duke University.

Following informed consent, participants answered the same questions used in the preliminary survey in order to create an ingroup and outgroup based on similarity between the current participant's and Mturk participants' responses. Using a computer algorithm, ingroup members were defined as the Mturk participants whose responses were most matched to the current participant's responses (with a minimum of at least 3 out of 5 questions matching for the political group and at least 4 out of 7 questions for the minimal group). Outgroup members were defined as the Mturk participants whose responses were least matched to the current participant's responses.

Dictator game data were collected using E-Prime 2.0 software (Psychology Software Tools, Inc., Pittsburgh, PA). All sessions began with a non-group control condition in which participants completed a series of one-shot dictator games (see Figure 2a). Participants were informed that the other person was a real person who agreed to be part of the study and that they could not allocate more points than were shown on the screen.

Following 30 control trials, 240 experimental trials were presented within a nested, blocked design. Subjects were randomly assigned to receive either the political group or minimal group manipulations first; then, within each group manipulation, they were randomly assigned to receive either the categorical or humanized conditions first. Finally, within each perception condition, ingroup and outgroup trials were intermixed. The trial structure was identical to the nongroup control, with the exception that group information (own/other) and one line of additional information (categorical/humanizing) were also presented (see Figure 2b). In the categorical condition, the additional information consisted of the person's response to one of the political (or art) preference questions. Participants were explicitly informed that because these responses were used to form the groups, this categorical information was redundant with the group membership information; that is, ingroup members' responses always matched the participant's response and outgroup members' responses always mismatched their response. In the humanized condition, the additional information consisted of the group member's response to one of the personal preference (icebreaker) questions. Participants were reminded that the group membership information (own/other) was based on political (or art) preferences even when viewing responses to the personal preference questions.

At the end of the experiment, one trial from each of the three group conditions (control, political, minimal) was randomly chosen for additional payment, using a conversion factor that participants were unaware of until the end of the study (100 points = $1). Additional survey measures were collected at the end of the study but are not included in the presented analyses. These included a modified recognition confusion task (44), need for cognition survey (45), universalism survey (46), social dominance survey (47), and personal need for structure survey (48).

Dictator Game: Data analysis. Data were analyzed using SPSS version 21. For each trial, we calculated the percentage of points allocated to the other person. Trials with allocations greater than 100% were excluded from analysis (0.71% of the trials). Averages were then calculated and entered into a 2 (group: ingroup/outgroup) x 2 (perception: categorical/humanized) repeated measures ANOVA. T-tests were conducted following significant interactions. The data fit a normal distribution (political groups: skew = 0.72, kurtosis = -0.19; minimal groups: skew = 0.58, kurtosis = -0.28), justifying the use of these tests. Partial η^2 values are included as measures of effect size.

We additionally created an ingroup bias score for each participant, calculated as a difference score (ingroup minus outgroup) of the percent transferred. This score measures the extent to which participants favored members of the ingroup over the outgroup when allocating points and was computed separately for categorical and humanized conditions. The comparison of these bias scores in the categorical and humanized conditions shows how the group bias changed when perception changed from categorical to humanized (see Figure 4).

Study 2 Methods

Participants. Before data collection, we established a target sample size of 76 participants (i.e., matching the sample size from Experiment 1). Data were collected from 83 participants recruited from Duke University and the surrounding community area. Using the same exclusion criteria as study 1, seven participants were excluded**.** The final data analysis consisted of 76 participants. The average age of the sample was 26.92 years old (SD = 11.52) and 67% were female.

Experimental Procedures. Procedures for study 2 were identical to study 1 with one exception: At the end of the experiment, participants rated each statement for how similar the statement was to themselves on a scale from 0 (not at all like me) to 100 (exactly like me). Responses were recorded using a slider bar presented in E-prime. Data analysis consisted of the same series of repeated measures ANOVAs and paired t-tests. Again, the data fit a normal distribution (political groups: skew = 0.68, kurtosis = 0.10; minimal groups: skew = 0.53, kurtosis = -0.02), justifying the use of these tests.

Similarity analysis. Analyses of similarity were restricted to the political group manipulation based on Study 1 results—there was a much larger dynamic range in that

manipulation, making it possible to test whether similarity alters pro-social behavior across groups. To account for the structure of our data, which contained trial-level similarity ratings for each subject, we used multilevel modeling with maximum likelihood estimation (49). Trials with allocations greater than 100% were excluded from analysis (0.29% of the trials) resulting in approximately 120 trials per subject. Subject was entered as a random factor and all other predictors (group and similarity) were entered as fixed factors.

To examine whether similarity ratings explained additional variance in allocation percentages, above and beyond the effects of group, we ran three nested models: 1) a baseline model with subject entered as a random factor, 2) a model with just subject and group, and 3) a model with subject, group, and similarity. This analysis was repeated for the categorical and humanized conditions, independently. To compare model fit, we conducted a likelihood ratio test using the -2 log likelihood between models.

In addition, we quantified the proportion of residual variance explained in each model, using an "individual-level variance explained" measure, calculated as $1 - \sigma^2_{\text{model x}} / \sigma^2_{\text{model 1}}$ (see (49), p. 398-399). Importantly, this proportion is indexed relative to the *residual* variance in allocation amounts after accounting for differences between subjects and not the *total* variance in allocation amounts. We then calculated the change in residual variance ($\Delta \sigma^2$) at each additional step (see Table 2), which allows us to quantify the proportion of residual variance explained by adding each additional factor (group/similarity) to the model. This approach is similar to R^2 change in hierarchical OLS regression.

Groupiness rating. To examine whether the humanizing statements encouraged a second group identity, we collected data from an independent sample on the extent to which each statement prompted a group versus an individual identity. There were 336 unique statements seen in the dictator game (across all participants, all trials). To limit time requirements, three samples of participants were recruited from Amazon's Mechanical Turk (MTurk) to rate 112 statements each. Each sample consisted of 75-79 participants that completed the survey ($n_{\text{sample1}} = 75$; $n_{\text{sample2}} = 78$; $n_{\text{sample3}} = 79$). Participants were paid $0.67 for completing the ten-minute survey.

Following informed consent, MTurk participants were asked to rate each statement for the extent to which it made them think of the person providing the statement as a member of a group versus an individual. Participants made their response using a slider bar, with the labels "group" and "individual" at each end of the slider bar. The labels were counterbalanced and participants were explicitly told to pay attention to the ordering of the labels across trials. Importantly, the slider bar did not reveal the scale score associated with the participant's response while participants were completing the survey. During data analysis, all statement ratings (ranging from 0 to 100) were coded such that lower numbers represent a more group-like identity and higher numbers represent a more individuating identity. For each statement, an average groupiness rating was computed across raters.

Using multilevel modeling, we then examined whether the group membership effect (own/other) remained significant when controlling for the groupiness and similarity (and their interaction) of the humanizing statement. To reduce multicollinearity, similarity and groupiness were mean centered and the interaction term was created using the mean centered variables. Subject was entered as a random factor and all other predictors (group, similarity, groupiness, similarity x groupiness interaction) were entered as fixed factors. Upon finding a significant similarity x groupiness interaction for the minimal groups, we probed the interaction using simple slopes analysis at high and low levels of groupiness (2 standard deviations above/below the mean to match the approximate range of groupiness ratings).

References

1. Weinstein, N., & Ryan, R. M. (2010). When helping helps: autonomous motivation for prosocial behavior and its influence on well-being for the helper and recipient. *Journal of personality and social psychology*, *98*(2), 222.

2. Penner, L. A., Dovidio, J. F., Piliavin, J. A., & Schroeder, D. A. (2005). Prosocial behavior: Multilevel perspectives. *Annu. Rev. Psychol.*, *56*, 365-392.

3. Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (2010). Prejudice, stereotyping and discrimination: theoretical and empirical overview. *The SAGE handbook of prejudice, stereotyping and discrimination*, 3-29.

4. Majdandžić, J., Bauer, H., Windischberger, C., Moser, E., Engl, E., & Lamm, C. (2012). The human factor: Behavioral and neural correlates of humanized perception in moral decision making. *PLoS One*, *7*(10), e47698.

5. Morishima, Y., Schunk, D., Bruhin, A., Ruff, C. C., & Fehr, E. (2012). Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron*, *75*(1), 73-79.

6. Strombach, T., Weber, B., Hangebrauk, Z., Kenning, P., Karipidis, I. I., Tobler, P. N., & Kalenscher, T. (2015). Social discounting involves modulation of neural value signals by temporoparietal junction. *Proceedings of the National Academy of Sciences*, *112*(5), 1619-1624.

7. Imuta, K., Henry, J. D., Slaughter, V., Selcuk, B., & Ruffman, T. (2016). Theory of mind and prosocial behavior in childhood: A meta-analytic review.

8. Tusche, A., Böckler, A., Kanske, P., Trautwein, F. M., & Singer, T. (2016). Decoding the Charitable Brain: Empathy, Perspective Taking, and Attention Shifts Differentially Predict Altruistic Giving. *Journal of Neuroscience*, *36*(17), 4719-4732.

9. Harris, L. T., & Fiske, S. T. (2009). Social neuroscience evidence for dehumanised perception. *European review of social psychology*, *20*(1), 192-231.

10. Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of personality and social psychology*, *78*(4), 708.

11. Cuddy, A. J., Rock, M. S., & Norton, M. I. (2007). Aid in the aftermath of Hurricane Katrina: Inferences of secondary emotions and intergroup helping. *Group Processes & Intergroup Relations*, *10*(1), 107-118.

12. Masten, C. L., Morelli, S. A., & Eisenberger, N. I. (2011). An fMRI investigation of empathy for 'social pain'and subsequent prosocial behavior. *Neuroimage*, *55*(1), 381-388.

13. Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the low: Neuroimaging responses to extreme outgroups. *Psychological Science, 17*, 847–853.

14. Leyens, J. P., Rodriguez-Perez, A., Rodriguez-Torres, R., Gaunt, R., Paladino, M. P., Vaes, J., & Demoulin, S. (2001). Psychological essentialism and the differential attribution of uniquely human emotions to ingroups and outgroups. *European Journal of Social Psychology*, *31*(4), 395-411.

15. Leyens, J. P., Cortes, B., Demoulin, S., Dovidio, J. F., Fiske, S. T., Gaunt, R., ... & Vaes, J. (2003). Emotional prejudice, essentialism, and nationalism The 2002 Tajfel Lecture. *European Journal of Social Psychology*, *33*(6), 703-717.

16. Cuddy, A. J., Fiske, S. T., & Glick, P. (2007). The BIAS map: behaviors from intergroup affect and stereotypes. *Journal of personality and social psychology*, *92*(4), 631.

17. 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*(3), 573-584.

18. Vaes, J., Leyens, J. P., Paola Paladino, M., & Pires Miranda, M. (2012). We are human, they are not: Driving forces behind outgroup dehumanisation and the humanisation of the ingroup. *European Review of Social Psychology*, *23*(1), 64-106.

19. Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2008). The neural substrates of in-group bias: A functional magnetic resonance imaging investigation. Psychological Science, 19, 1131–1139.

20. Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2011). Modulation of the Fusiform Face Area following minimal exposure to motivationally relevant faces: Evidence of ingroup enhancement (not out-group disregard). Journal of Cognitive Neuroscience, 23, 3343–3354.

21. Volz, K. G., Kessler, T., & von Cramon, D. Y. (2009). In-group as part of the self: In-group favoritism is mediated by medial prefrontal cortex activation. Social Neuroscience, 4, 244–260.

22. Falk, E. B., Spunt, R. P., & Lieberman, M. D. (2012). Ascribing beliefs to ingroup and outgroup political candidates: neural correlates of perspective-taking, issue importance and days until the election. *Phil. Trans. R. Soc. B*, *367*(1589), 731-743.

23. Molenberghs, P., & Morrison, S. (2014). The role of the medial prefrontal cortex in social categorization. *Social cognitive and affective neuroscience*,*9*(3), 292-296.

24. Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. *The social psychology of intergroup relations*, *33*(47), 74.

25. Levine, M., Prosser, A., Evans, D., & Reicher, S. (2005). Identity and emergency intervention: How social group membership and inclusiveness of group boundaries shape helping behavior. *Personality and Social Psychology Bulletin*, *31*(4), 443-453.

26. Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European journal of social psychology*, *1*(2), 149-178.

27. Harris, L. T., & Fiske, S. T. (2007). Social groups that elicit disgust are differentially processed in the mPFC. *Social Cognitive Affective Neuroscience, 2,* 45–51.

28. Wheeler, M. E., & Fiske, S. T. (2005). Controlling racial prejudice social-cognitive goals affect amygdala and stereotype activation. *Psychological Science*, *16*(1), 56-63.

29. Baumgartner, T., Götte, L., Gügler, R., & Fehr, E. (2012). The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. *Human brain mapping*, *33*(6), 1452-1469.

30. Baumgartner, T., Nash, K., Hill, C., & Knoch, D. (2015). Neuroanatomy of intergroup bias: A white matter microstructure study of individual differences.*NeuroImage*, *122*, 345-354.

31. Bruneau, E. G., Cikara, M., & Saxe, R. (2015). Minding the Gap: Narrative Descriptions about Mental States Attenuate Parochial Empathy. *PloS one*, *10*(10), e0140838.

32. Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619-619.

33. Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in cognitive sciences*, *14*(8), 383-388.

34. Parker, M. T., & Janoff-Bulman, R. (2013). Lessons from morality-based social identity: The power of outgroup "hate," not just ingroup "love". *Social Justice Research*, *26*(1), 81-96.

35. Jenke, L., & Huettel, S. A. (2016). Issues or Identity? Cognitive Foundations of Voter Choice. *Trends in Cognitive Sciences*, *20*(11), 794-804.

36. Rachlin, H., & Jones, B. A. (2008). Altruism among relatives and non-relatives. *Behavioural processes*, *79*(2), 120-123.

37. Goeree, J. K., McConnell, M. A., Mitchell, T., Tromp, T., & Yariv, L. (2010). The 1/d law of giving. *American Economic Journal: Microeconomics*, *2*(1), 183-203.

38. Lee, V.K. & Harris, L.T. (2013) Dehumanized Perception: Psychological and Neural Mechanisms Underlying Everyday Dehumanization. In P.G. Bain, J. Vaes, & J-Ph. Leyens (Eds).*Humanness and dehumanization*. Psychology Press.

39. Haque, O. S., & Waytz, A. (2012). Dehumanization in medicine causes, solutions, and functions. *Perspectives on Psychological Science*, *7*(2), 176-186.

40. Harris, L. T., Lee, V. K., Capestany, B. H., Cohen, A. O. (2014). Assigning economic value to people results in dehumanization brain response. Journal of Neuroscience, Psychology, and Economics.

41. Cameron, C. D., & Payne, B. K. (2011). Escaping affect: how motivated emotion regulation creates insensitivity to mass suffering. *Journal of personality and social psychology*, *100*(1), 1.

42. Genevsky, A., Västfjäll, D., Slovic, P., & Knutson, B. (2013). Neural underpinnings of the identifiable victim effect: Affect shifts preferences for giving. *The Journal of Neuroscience*, *33*(43), 17188-17196.

43. Gaertner, S. L., & Dovidio, J. F. (2014). *Reducing intergroup bias: The common ingroup identity model*. Psychology Press.

44. Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, *36*, 778–793.

45. Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological bulletin*, *119*(2), 197.

46. Phillips, S. T., & Ziller, R. C. (1997). Toward a theory and measure of the nature of nonprejudice. *Journal of Personality and Social Psychology*, *72*(2), 420.

47. Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of personality and social psychology*, *67*(4), 741.

48. Thompson, M. M., Naccarato, M. E., & Parker, K. E. (1989, June). Assessing cognitive need: The development of the Personal Need for Structure and Personal Fear of

Invalidity scales. Paper presented at the annual meeting of the Canadian Psychological Association, Halifax, Nova Scotia.

49. Hayes, A. F. (2006). A primer on multilevel modeling. *Human Communication Research*, *32*(4), 385-410.

Figure Legends

Figure 1: Predicted results for three potential models of humanization. Compared to traditional categorical group membership and a nongroup control (solid bars), three distinct models make predictions for how humanization affects group members (shaded bars). All three models predict reduced group bias and an increase in prosocial behavior for humanized outgroup members, but differ in predictions made for humanized ingroup members. Humanizing ingroup members may increase (generalized humanization model), produce no change (outgroup humanization model), or decrease (competitive humanization model) prosocial behavior. Note: arrows predict the directional effect of humanization and not necessarily the magnitude of the shift in prosocial behavior.

Figure 2: Study design. a) All study sessions began with a nongroup control condition in which participants were given a sum of points and asked to indicate how many points (if any) they would allocate to another person. Participants' responses were entered using the number keys on the keyboard and displayed on the screen. After entering a response (self-paced), the trial proceeded to an inter-trial fixation (2 sec) and onto the next trial. b) Example trials from our 2x2 study design in which participants were provided group membership and person perception information.

Figure 3: Allocation Behavior in Control and Political Groups. Bar graphs depicting participants' prosocial behavior in the dictator game in a) Study 1 ($n = 76$) and b) Study 2 ($n = 76$). Error bars represent the standard error of the mean. * indicates significant differences at $p < 0.001$.

Figure 4: Individuation consistently reduces ingroup biases associated with political groups. An ingroup bias score was calculated for each participant by taking a difference score in the percent of money transferred to ingroup and outgroup members. Larger numbers along the y-axis represent greater allocations to ingroup than outgroup members. Each participant is represented on the x-axis by two consecutive bars: one for the ingroup

bias under categorical perception (blue), and the other for the ingroup bias under individuated perception (orange). For nearly all participants who showed an ingroup bias under categorical perception, their ingroup bias was reduced or eliminated under individuated perception. Studies 1 and 2 are represented together, with participants ordered according to (diminishing) ingroup bias in the categorical condition.

Figure 1

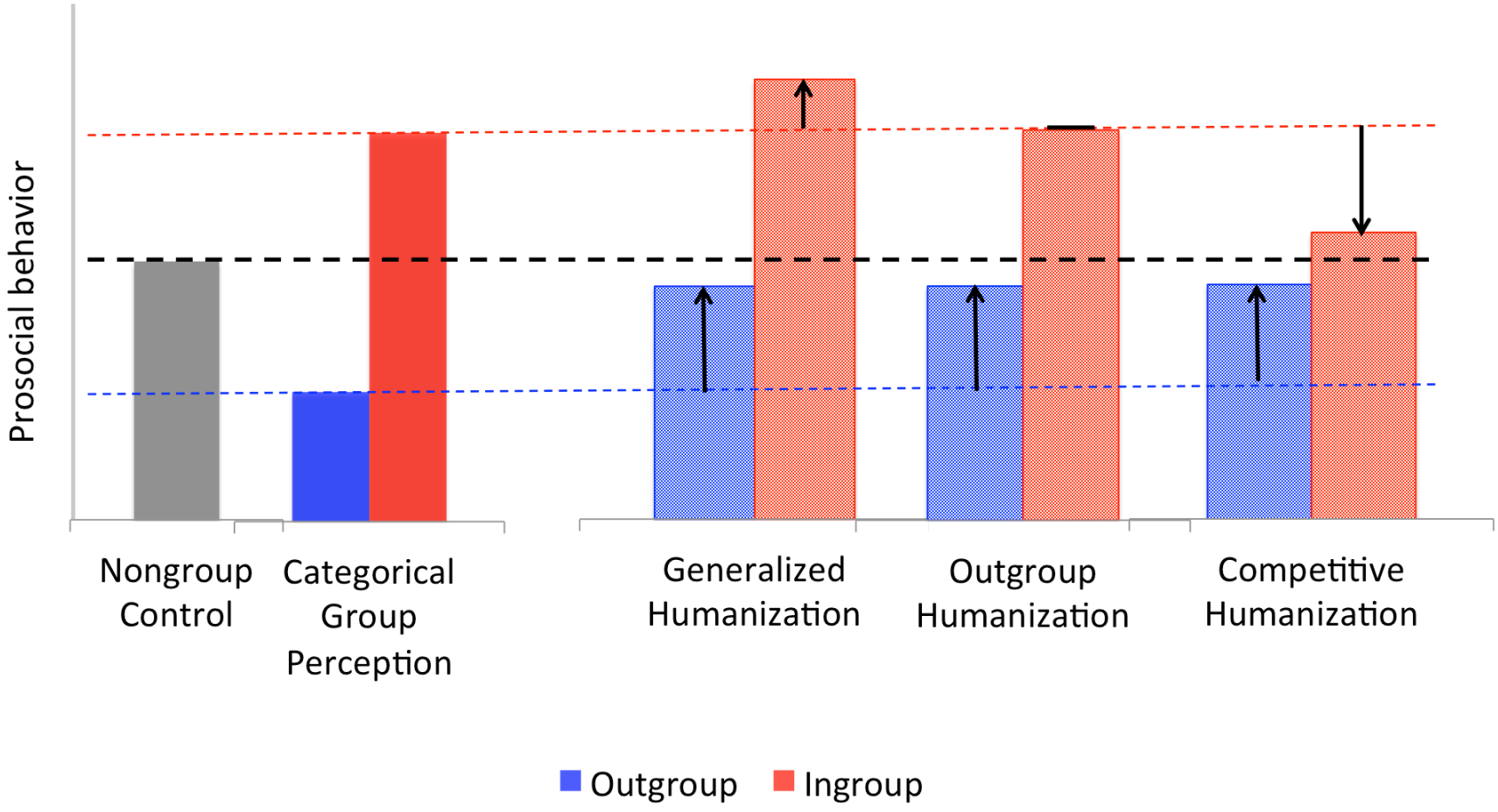
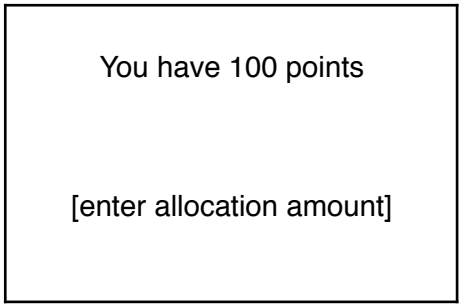

Prosocial behavior
Nongroup Control
Categorical Group Perception
Generalized Humanization
Outgroup Humanization
Competitive Humanization
Outgroup
Ingroup

Figure 2

a Nongroup Control



b Group manipulation

	Ingroup	Outgroup
Categorical	You have 100 points Group: own I think gay marriage should be legally recognized	You have 100 points Group: other I think gay marriage should not be recognized
Humanized	You have 100 points Group: own When I was little I wanted to be a teacher when I grew up	You have 100 points Group: other When I was little I wanted to be a doctor when I grew up

Figure 3

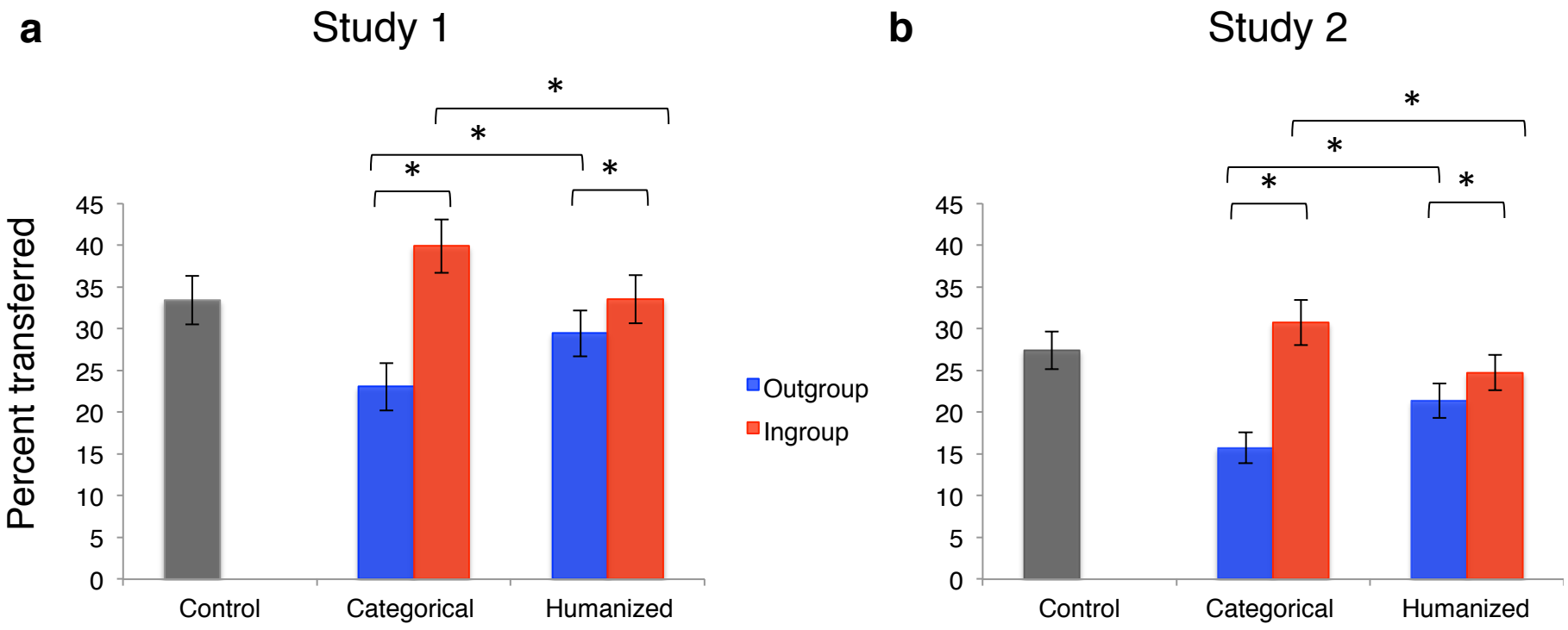

a
Study 1
Percent transferred
45
40
35
30
25
20
15
10
5
0
Control
Categorical
Humanized
*
*
*
*
Outgroup
Ingroup
b
Study 2
45
40
35
30
25
20
15
10
5
0
Control
Categorical
Humanized
*
*
*
*

Figure 4

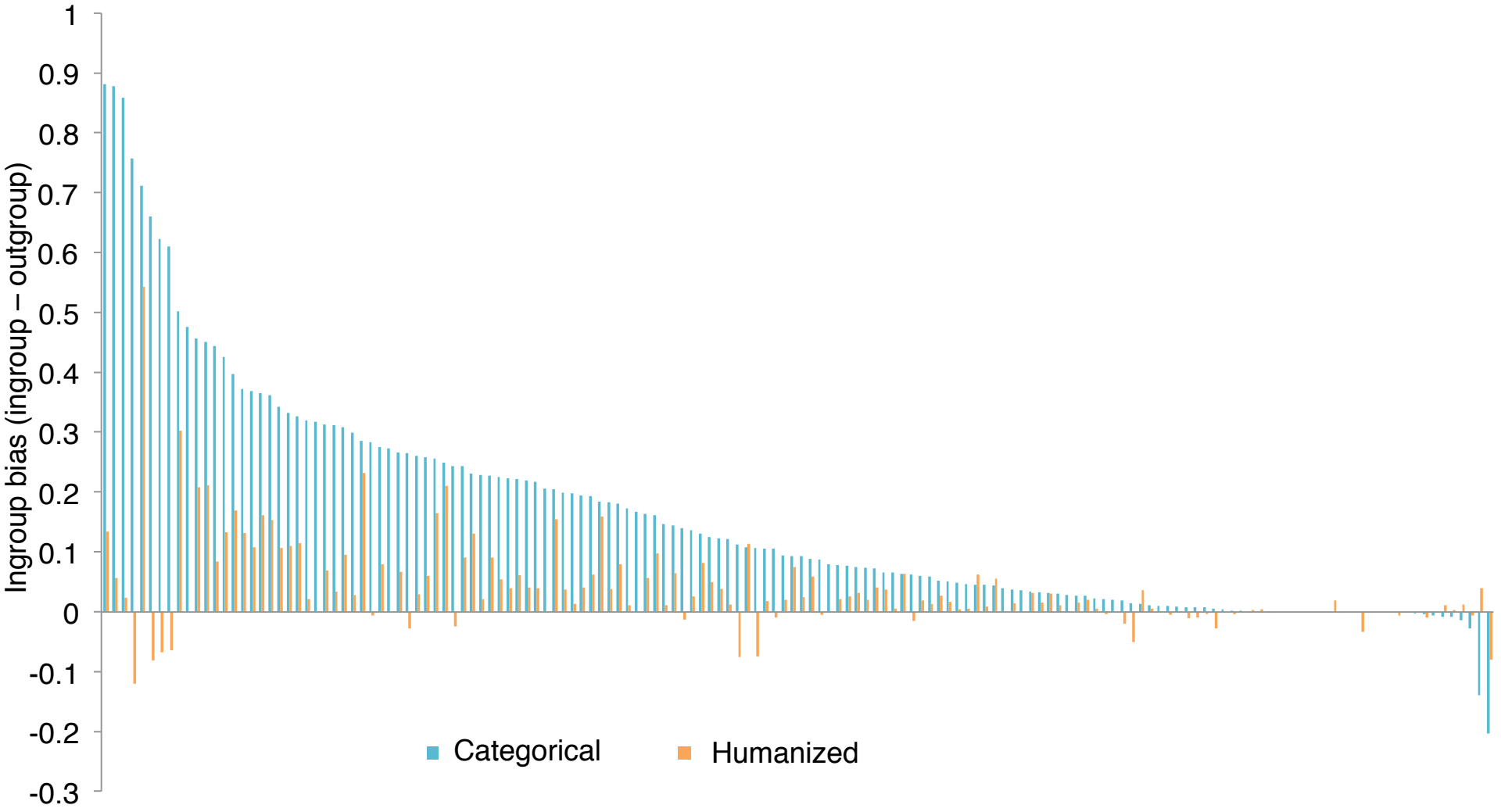

1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
Ingroup bias (ingroup – outgroup)
Categorical
Humanized

Table 1. Behavioral results from Study 1 and Study 2: Average allocations in the dictator game (standard deviation in parentheses)

	Categorical	Humanized
Study 1		
Control	33% (25%)	
Political Outgroup	23% (25%)	29% (24%)
Political Ingroup	40% (28%)	34% (25%)
Minimal Outgroup	30% (25%)	31% (24%)
Minimal Ingroup	36% (27%)	32% (25%)
Study 2		
Control	27% (20%)	
Political Outgroup	16% (16%)	21% (18%)
Political Ingroup	31% (23%)	25% (19%)
Minimal Outgroup	23% (20%)	24% (18%)
Minimal Ingroup	27% (21%)	25% (18%)

Table 2. Parameter estimates for nested models examining the effect of similarity for categorical and humanized perception.

Condition	ICC	Model	Deviance (-2LL)	Likelihood ratio test X^2 (df = 1) all $p < .001$	τ_{00}	σ^2 (95% CI)	Δ residual variance ($\Delta \sigma^2$)
Categorical	0.5955	1	31320.82		319.42	216.96 (207.31, 227.06)	
		2	30301.75	1019.07	320.29	164.90 (157.57, 172.57)	-52.06 (24%)
		3	30236.63	65.12	322.48	162.01 (154.81, 169.55)	-2.88 (1.3%)
Humanized	0.6800	1	35907.55		317.76	149.51 (143.43, 155.85)	
		2	35820.50	87.05	317.91	146.62 (140.65, 152.84)	-2.89 (1.9%)
		3	35414.61	405.89	311.43	133.90 (128.45, 139.57)	-12.72 (8.5%)