Note: The present paper is in press at the *Journal of Experimental Social Psychology.*

Simulation: A Strategy for Mindreading Similar but Not Dissimilar Others?

Brandon M. Woo and Jason P. Mitchell

Harvard University

Author Note

Brandon M. Woo, Department of Psychology, Harvard University.

Correspondence concerning this article should be addressed to Brandon M. Woo, Department of Psychology, Harvard University, Cambridge, MA. E-mail: bmwoo@g.harvard.edu.

Abstract

One proposed solution to the challenge of mentalizing is to use one's own mind as a model in a process known as simulation. Studies have demonstrated that people more strongly engage in simulation for individuals to whom they feel more similar, such that people assume that the self is like more similar others (and unlike more dissimilar others) even in attitudes that are irrelevant to reasons for perceiving similarity (e.g., group membership). The present 3 studies (total $N = 193$) evaluate this strategy using three kinds of groups, both real and novel, that participants may feel differentially similar about. These studies conceptually replicate past work, finding that people simulate when reasoning about more similar others, despite the attitudes in question being irrelevant to group membership. Additionally, these studies find that the more similar people feel to a target, the more correct people are about the attitudes of individuals in a target's group. These findings suggest that although people may simulate more strongly for more similar others even when attitudes are irrelevant to group membership, it may be wrong to think that the self is unlike more dissimilar others. Finally, the present studies provide preliminary evidence that when perceivers' attitudes are unlike most others' attitudes, such that simulation could be inappropriate, perceivers do not appropriately correct the extent to which they simulate for more similar others. These results suggest that there are limitations to the strategy of selectively simulating for more similar others in order to understand others' minds.

Keywords: mentalizing; social cognition; simulation; self

Simulation: A Strategy for Mindreading Similar but Not Dissimilar Others?

In everyday social interaction, people face the challenge of making sense of the mental states—the beliefs, desires, attitudes, etc.—of other people despite lacking direct access to others' minds. One proposed solution to this challenge of understanding others' minds is to use the one mind to which people do have direct access—their own—as a starting point in a process known as simulation (Davies & Stone, 1995a, 1995b; Epley & Waytz, 2009; Gallese & Goldman, 1998; Heal, 1986; Meltzoff & Brooks, 2001; Nickerson, 1999; Van Veelen, Otten, Cadinu, & Hansen, 2016). A host of research has demonstrated that people may infer others' mental states by starting with their own and projecting their own mental states onto others, assuming that others know what they know, see what they see, and like what they like (Camerer, Loewenstein, & Weber, 1989; Keysar, Lin, & Barr, 2003; Nickerson, 1999; Van Boven, Dunning, and Loewenstein, 2000; Van Veelen et al., 2016; for review of complementary neuroimaging evidence, see Mitchell, Banaji, & Macrae, 2005; Waytz & Mitchell, 2011; cf. Saxe, 2005). In the present paper, we examine the appropriateness of this strategy when reasoning about others' attitudes.

The Selective Simulation of More Similar Others

Although some research has framed simulation as a process for how people reason about others' minds generally, the structure of the social world may have profound implications for simulation. People readily organize themselves into social groups based on to whom they feel similar (e.g., because of shared cues such as political orientation, language, ethnicity, etc.), distinguishing between groups to whom they feel more or less similar (see Dunham, 2018; Turner, 1985; Turner & Oakes, 1986, 1989). This sensitivity to similarity is present early in childhood, such that infants distinguish between someone who speaks in their native language

versus someone who speaks in a foreign language (Kinzler, Dupoux, & Spelke, 2007), and expect two people who speak in the same language to affiliate with each other (Liberman, Woodward, & Kinzler, 2017). Moreover, this sensitivity to similarity is not limited to pre-existing cues of similarity like political orientation or language that people have likely had years of experience identifying with; beginning with Tajfel (1970), studies have demonstrated that people even prefer someone who is like them based on arbitrary, novel cues, such as shared outcomes on a coin-flip (see Dunham, 2018; Otten, 2016). Taken together, this body of research demonstrates that social groups and perceived similarity are fundamental to human psychology.

Critically, it may make less sense to simulate others' mental states if the target of mindreading is too socially distant (e.g., belonging to another social group, which may follow different principles). The self may appear to be a better model for understanding another person's attitudes the more similar the self is to another person (Ames, 2004a, 2004b; Mitchell, Banaji, & Macrae, 2005). Being in the same social group may provide reasons to believe that the self is similar to another person, such that people more readily engage in simulation. Conversely, being in a different social group may provide reasons to believe that the self is unlike another person, such that people do not readily engage in simulation (see Turner, 1985; Turner & Oakes, 1986, 1989). Indeed, studies have demonstrated that simulation occurs more strongly when reasoning about the attitudes of ingroup members than when reasoning about the attitudes of outgroup members (for review and meta-analysis, see Robins & Krueger, 2005).

Increased simulation for ingroup versus outgroup members may seem obvious when mental states are relevant to the group's nature. For example, if one is in a political group, others in the group may be likely to share one's politically relevant attitudes (Chambers, Baron, & Inman, 2006). Strikingly, the tendency to more strongly simulate for ingroup members is

intrusive, occurring even when mental states are irrelevant to the group's nature, such that there is great correspondence between people's reports of their own mental states and the mental states of groups that people feel similar to. For instance, Clement and Krueger (2002) found that after sorting people into two groups, ostensibly based on how people responded to a geometric task, people thought that someone who responded to the geometric task similarly to themselves would also be more like themselves on unrelated personality items. Similarly, Tamir and Mitchell (2013) found that people assume that someone of the same political orientation (a political ingroup member) is more like themselves in apolitical attitudes (e.g., enjoyment of exercise) than a political outgroup member. Complementary neuroimaging work suggests that the same neural machinery underlies the consideration of one's own apolitical attitudes and the apolitical attitudes of political ingroup, but not outgroup, members (Jenkins, Macrae, & Mitchell, 2008; Mitchell, Macrae, & Banaji, 2006).

Importantly, these findings of selective simulation of more similar others cannot be reduced to: (i) people having stereotypes about groups that may influence how they reason about the members of different groups given that selective simulation of similar others occurs even when attitudes are irrelevant to the group (see Ames, Mor, & Toma, 2013; Cho & Knowles, 2013; Epley & Waytz, 2009; Greenwald et al., 2002; Van Veelen et al., 2016); or (ii) people systematically attributing fewer or weaker levels of mental states to dissimilar others (for evidence that people may dehumanize such groups, see Harris & Fiske, 2006; O'Brien & Ellsworth, 2012; Xu, Zuo, Wang, & Han, 2009). Rather, people are perceiving the mental states of members of groups that they feel dissimilar to as being dissimilar to their own. Furthermore, to the extent that people actively engage in perspective-taking for someone of the opposite political orientation, which may make people feel more like someone who is politically

dissimilar to themselves, people assume that their apolitical attitudes are more like those of politically dissimilar others; and to the extent that people actively engage in perspective-taking for someone of the same political orientation, which may make people appreciate differences between the self and politically similar others, people assume that their apolitical attitudes are less like those of politically similar others (Todd, Simpson, & Tamir, 2016). In sum, then, a large body of literature has characterized the use of the self as a model for understanding the attitudes of similar but not dissimilar others, even in circumstances in which such simulation may be unwarranted.

The Correctness of Selectively Simulating More Similar Others

Beginning with early psychologists like Ichheiser (1949) and Cronbach (1955), studies have examined the accuracy of person perception (see Kenny & Albright, 1987; Swann, 1984). The present studies tackle the question of whether the use of the self as a model for understanding the attitudes of similar but not dissimilar others is an appropriate strategy. That is, do perceivers get any closer to the truth of another's attitudes for simulating more for individuals whom they feel more similar to and simulating less (if at all) for individuals whom they feel more dissimilar to? Does the selective simulation of more similar others lead people to be more correct about others' attitudes?

The selective simulation of more similar others could only be appropriate if there are few differences between the self and more similar others, large differences between the self and more dissimilar others, and consequently, large differences between similar others and dissimilar others. There are multiple ways in which perceivers could be wrong for differentially simulating based on perceived similarity. Perceivers may be erroneously assuming: similarity in attitudes based on sharing qualities (e.g., group membership) that are irrelevant to the attitudes;

dissimilarity in attitudes based on not sharing qualities that are irrelevant; or both. Such errors could have implications for interactions both within and between groups. The implications for intergroup relations could be particularly important. A number of studies have demonstrated that people like others more if they share more attitudes (Byrne, 1969, 1971; Byrne & Nelson, 1965; Kaplan & Anderson, 1973; Rosenbaum, 1986; Singh & Ho, 2000). If people are seeing unfounded differences in attitudes between themselves and individuals from other groups, then people might also like individuals from other groups less and be less likely to cooperate with individuals from other groups.

There is some evidence that people may be wrong to selectively simulate only for more similar others. Hoch (1987) found that graduate students are more correct about the attitudes of their spouse, their peers in graduate school, and typical married, American consumers on a wide range of topics (e.g., social issues, finance, and food) if they simulated more. Similarly, Mor, Toma, Schweinsberg, and Ames (2019) found that the more people from individualistic cultures assume that people from China are like themselves in individualistic attitudes, the more correct people from individualistic cultures were about the individualistic values of people from China today. Although this evidence suggests that people are more correct to simulate generally, regardless of the group affiliation of the target of simulation, importantly, this past work has at least two critical limitations.

First, past work has not explored how correct people would be to selectively simulate more similar others when attitudes are irrelevant to the shared qualities (e.g., group membership). Yet, other research has demonstrated that people more strongly simulate similar others' attitudes, even when attitudes are irrelevant to reasons for perceiving similarity (e.g., Clement & Krueger, 2002; Tamir & Mitchell, 2013). It is unknown how correct people may be

to selectively simulate more similar others when the attitudes in question are irrelevant to group membership. If people engage in such selective simulation and are more correct to simulate, even when attitudes in question are irrelevant to group membership, then people should be least correct about the attitudes of more dissimilar others. Such a pattern of findings would suggest that people *should* simulate for more dissimilar others.

Second, past work on the correctness of simulation has neglected the range of individual differences in people's attitudes. That is, it may only be correct to use one's own mind as a starting point for reasoning about others' attitudes if one's own attitudes are typical. If one's attitudes are atypical, and one still simulates when reasoning about others' attitudes, then one would be wrong about others' attitudes. It is unknown if perceivers recognize when their attitudes are atypical and simulate less in such circumstances.

The Present Studies

In the present set of studies, we first replicated the finding that people more strongly simulate the attitudes of individuals that they perceive to be more similar, despite such attitudes being seemingly irrelevant to reasons (e.g., group membership) for perceiving individuals to be differentially similar. This replication enabled us to examine our central question of whether correctness in perceivers' ratings of others' attitudes varies depending on perceived similarity, even when the attitudes in question are irrelevant to group membership, the main reason for perceiving similarity in the present studies. Specifically, we examined whether perceivers are correct to simulate more strongly for more similar versus less similar others in contexts in which group membership is unrelated to attitudes. Additionally, we conducted exploratory analyses to investigate whether perceivers correct the degree to which they simulate when their attitudes are atypical.

In three studies, we introduced participants to an individual representative of one group to which approximately half of the participants belonged and an individual representative of a second group to which the remaining participants belonged. In describing the group membership of the two individuals, we provided a reason for participants to feel differentially similar to the two individuals. After participants learned of the two individuals, we asked participants to rate how well 100 statements applied to themselves and to both of the individuals. These ratings enabled us to understand participants' attitudes (i.e., the ground truth for the two groups of participants), and what they thought might be the attitudes that someone in each group holds towards these statements. Across the statistical analyses that we conducted for the replication and to address our central question, we examined the predictive effects of both perceived similarity and group membership.

Across studies, we established group membership differently. Study 1 used political orientation, establishing groups that exist prominently in the real world. Study 2 divided participants into groups (Eagles/Rattlers) based on whether their birthdates were on odd- or even-numbered days, establishing groups that exist but are less prominent. We recognized that this was particularly minimal and odd in Study 2, given that the group distinction is highly uncommon in real life, but wanted to see whether participants would still engage in intrusive simulation with such minimal information (see Dunham, 2018). Study 3 used randomly assigned novel groups (Eagles/Rattlers), based ostensibly on personality tests, establishing groups for which there could not be prior knowledge about how the groups differed in attitudes. In using different ways to establish groups (pre-existing and novel groups; groups for which the distinction between ingroup and outgroup is more or less prominent), we can examine the generalizability of our findings to cases in which there is less of a basis for differentiating two

groups. Our studies shed light on what cues trigger the use of the self as a model for understanding others' minds.

Method

We report all measures, manipulations, and exclusions in these studies.

Participants

Participants were recruited via Amazon Mechanical Turk. Samples of 66, 66, and 61 participants were used in Study 1 (*M* age = 37.44 [*SD* = 11.59]; 32 women, 33 men, 1 other; 45 liberals, 21 conservatives), Study 2 (*M* age = 39.9 [*SD* = 9.86]; 44 women, 22 men; 33 Eagles, 33 Rattlers), and Study 3 (*M* age = 37.00 [*SD* = 11.90]; 34 women, 27 men; 30 Eagles, 31 Rattlers), respectively. An additional 27 participants, 25 participants, and 21 participants completed Studies 1, 2, and 3, respectively, but were excluded from analyses if they: failed an attention check (*ns* = 8, 1, and 5 in Studies 1, 2, and 3, respectively); or failed manipulation checks (*n*s = 19, 24, and 20 in Studies 1, 2, and 3, respectively).

Sample size justification. Sample size was based on power simulations on pilot data (*n* = 26) with models assessing whether the extent to which participants felt similar to a target representative of a group (liberals/conservatives, as in Study 1) predicted the extent to which participants were correct about the attitudes of the target's group linearly. With 60 subjects, $\alpha = .05$, and the assumption that $\beta = 0.14$ (as in the model based on pilot data), power was 1 within machine precision. This sample size was determined before collecting the data and conducting data analyses for the present (i.e., non-pilot) studies.

Procedure

General structure. In each study, participants first indicated their age, gender, and ethnicity. We then introduced participants to two different targets who belonged to one of two

social groups: liberals/conservatives in Study 1, Eagles/Rattlers in Studies 2 and 3. Most participants self-identified more strongly with the liberal or the conservative target in Study 1, and all participants were assigned to be an Eagle or Rattler in Studies 2 and 3. Critically, for Studies 2 and 3, one target was in the same social group (ingroup) as the participant and one target was in a different social group (outgroup). For Study 1, there were similarly ingroup and outgroup targets if participants self-identified differently with the liberal and the conservative targets.

Participants saw photographs of the two targets, who were gender-matched to participants. (If participants answered "other" for gender, they viewed male targets.) We matched photographs on age, attractiveness, and ethnicity (see Bainbridge, Isola, & Oliva, 2013, the source of the photographs), and randomized which of the two photographs participants saw for each target.

Participants completed a judgment task, in which they answered questions about their attitudes and those of each target. Participants were told that we were curious about how people make judgments about whether a statement applies to someone. On each trial, participants saw one of 100 statements (e.g., "enjoy exercising", "love to people watch", "enjoy singing in the shower", "think school is important"; for full list, see "Open Practices" section), and judged how well the statement applied to themselves (self-ratings) or one of the targets using a ten-point scale anchored on: 1 = extremely unlikely, 10 = extremely likely. Participants judged the self and each target in separate blocks in random order (i.e., one block each for the self and the two targets). Following the judgment task, participants used a ten-point scale to indicate how similar they felt to each target.

Group membership. In Study 1, targets differed by their political affiliation (liberal/conservative). Before the judgment task, participants saw photographs of two individuals, one of whom was described as a liberal, Democratic supporter of Hillary Clinton and one of whom was described as a conservative, Republican supporter of Donald Trump (see Mitchell et al., 2006; Tamir & Mitchell, 2013 for similar manipulations). Following the judgment task, participants used a ten-point scale to indicate how liberal/conservative each target was and how liberal/conservative they themselves were. Group membership of participants was determined by whether participants indicated that they were more liberal or more conservative.

In Study 2, targets differed by membership in one of two groups (Eagles/Rattlers) that were based on whether birthdates were odd-numbered or even-numbered. Similarly, we assigned participants to one of two novel groups on the basis of their birthdate. Participants with odd-numbered birthdates (e.g., February 15th) were Rattlers; participants with even-numbered birthdates (e.g., February 16th) were Eagles. We informed participants that everyone, including the targets, were assigned to groups following this rule.

In Study 3, targets differed by membership in one of two novel groups (Eagles/Rattlers) that were ostensibly based on differences in personality. Before being introduced to targets, participants answered 5 items from the Big 5 Personality Inventory (Goldberg, 1992). Participants were then (falsely) told that their scores indicated membership in one of two groups (Eagles/Rattlers), which comprised other people who completed the personality test similarly (as in Bruneau, Cikara, & Saxe, 2015). In fact, team membership was randomly assigned. Critically, participants were not given an explanation of what parts of the personality test differentiated Eagles from Rattlers.

Manipulation checks. We asked participants to indicate whether each target supported Clinton or Trump in Study 1, and we asked participants to indicate the group membership of each target and whether they themselves were an Eagle or Rattler in Studies 2 and 3.

Results

Differences Between Established Groups

Previous work has assumed that items like the present study's 100 items were apolitical. To verify this in Study 1 and ensure that items did not differ between groups in Studies 2 and 3, we conducted planned analyses to determine whether there were differences between groups for all studies. Out of 100 statements, liberals and conservative participants responded significantly (ps < .05) differently on 8 statements, and Eagles and Rattlers responded significantly differently on 4 statements in Study 2 and on 0 statements in Study 3. That is, we observed differences between groups on either no more or slightly more items than expected by chance (i.e., 5%).

Perceived Differences Between Groups

In exploratory analyses, we investigated whether there were items on which participants thought that the targets of the groups differed from each other in a consistent direction (e.g., consistently thinking that the liberal target was higher than a conservative target in an item X, or consistently thinking that the Eagle target was higher than a Rattler target in an item Y). Out of 100 statements, participants rated the targets as being different (ps < .05) from each other in a consistent direction on 11 statements in Study 1, 3 statements in Study 2, and on 10 statements in Study 3. The number of items on which participants perceived differences between targets in a

consistent direction, then, were either below or slightly above chance levels, and were comparable in studies involving real and novel groups (e.g., Study 1 versus Study 3).

Perceived Similarity to Each Group

We had provided group information of the two targets to provide a reason for participants to perceive similarity differentially between the two targets. In exploratory analyses, we aimed to verify whether group membership made a difference in perceived similarity. Across studies, participants perceived themselves to be more similar to target who was representative of the ingroup over the target who was representative of the outgroup (see Table 1 and Figure 1).

Table 1

Mean Perceived Similarity Scores for the Ingroup and Outgroup Target

	Ingroup	Outgroup	*t*	*df*s	*p*	*d*
Study 1	5.98 (2.08)	4.61 (2.14)	2.86	59	.005	0.64
Study 2	5.34 (2.04)	4.72 (2.15)	2.13	65	.036	0.29
Study 3	6.42 (2.31)	4.67 (2.35)	4.35	60	< .001	0.75

Note. Standard deviations are in brackets.

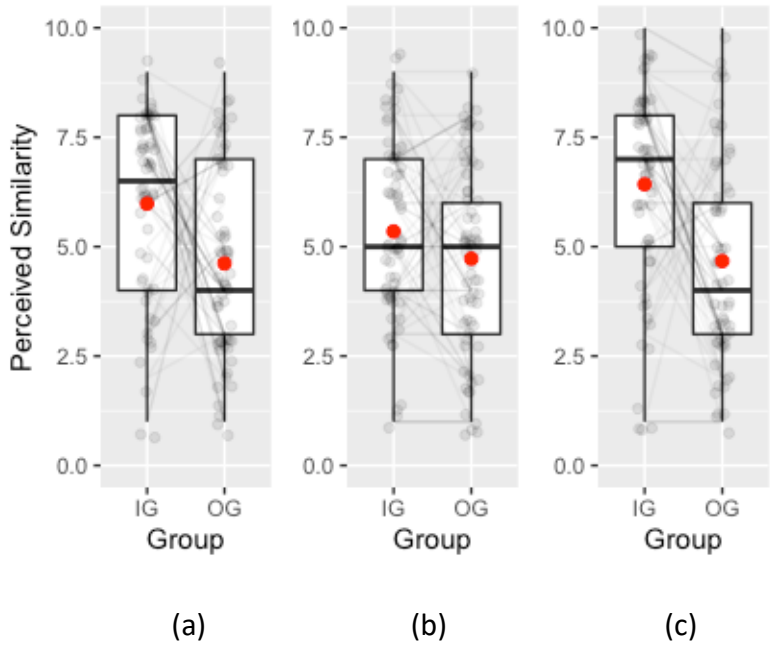


Figure 1. The relationship between group membership and perceived similarity scores in Studies 1 (a), 2 (b), and 3 (c). Red dots indicate means. Boxes indicate interquartile ranges and horizontal lines indicate medians. Translucent gray dots connected across boxes indicate similarity scores from individual participants. IG indicates ingroup and OG indicates outgroup.

Relation Between Self-Ratings and Ratings for a Target.

Having found that groups did not differ above chance in attitudes, that participants largely did not consistently think that the targets from the two groups differed in attitudes above chance, and that group membership influenced perceived similarity, we next conducted planned analyses to examine the relation between participants' own attitudes and their judgments of the attitudes of relatively more similar others. For each study, we ran a mixed-effects model in which the dependent variable was the rating for a target on a specific item; the fixed effects were the self-rating for the item, how similar participants felt to the target relative to how similar the participants felt to the other target (i.e., relative similarity; the difference of two ratings of perceived similarity), group membership (ingroup = -0.5, outgroup = 0.5), the interaction

between self-rating and relative similarity, and the interaction between self-rating and group membership; and the random effects were participant ID and item number. In the models, fixed effects were centered.

Across studies, the more participants perceived themselves to be relatively more similar to a target versus the other target, the more positively their self-ratings on attitudes predicted their ratings for that target (see Table 2 and Figure 2). Additionally, in Studies 1 and 3, but not in Study 2, participants' self-ratings on attitudes more strongly predicted their ratings of the ingroup target than those of the outgroup target (see Table 2 and Figure 3; for additional exploratory analyses on the relationships between similarity ratings, group membership, participants' own attitudes, and participants' judgments of the attitudes of a target, see Supplementary Material). Critically, across studies, the interaction involving perceived similarity was larger in effect size than the interaction involving group membership.

Table 2

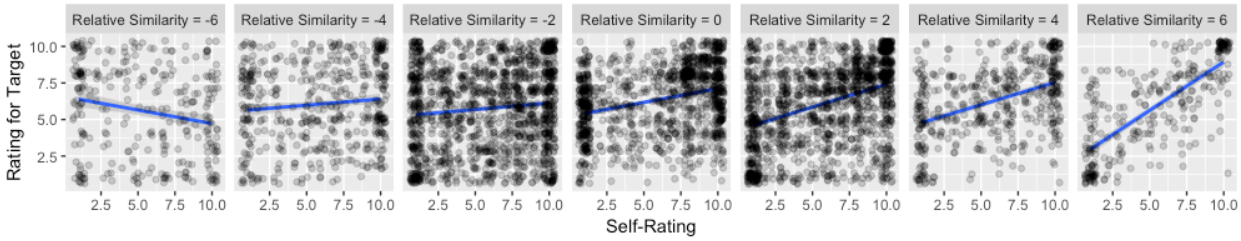
The Interactions Between Self-Ratings, Relative Similarity, and Group Membership in Predicting Ratings of a Target

	β	*t*	*df*s	*p*	95% CI
Study 1					
Self-Rating	0.17	19.26	12790	< .001	[0.15, 0.19]
Relative Similarity	0.05	7.19	13080	< .001	[0.04, 0.07]
Group Membership	-0.00	-0.55	13030	.583	[-0.01, 0.01]
Interaction between SR and RS	0.20	26.36	13030	< .001	[0.17, 0.20]
Interaction between SR and GM	0.04	5.26	13030	< .001	[0.02, 0.05]
Study 2					
Self-Rating	0.22	25.34	12760	< .001	[0.20, 0.24]
Relative Similarity	0.01	1.58	13030	.113	[-0.0002, 0.02]
Group Membership	-0.008	-1.07	13030	.284	[-0.02, 0.006]
Interaction between SR and RS	0.14	18.71	13030	< .001	[0.13, 0.16]
Interaction between SR and GM	-0.001	-0.19	13030	.843	[-0.01, 0.01]
Study 3					
Self-Rating	0.21	23.31	11360	< .001	[0.19, 0.22]
Relative Similarity	0.04	6.24	12030	< .001	[0.03, 0.06]
Group Membership	-0.00	36.22	10720	.964	[-0.01, 0.01]
Interaction between SR and RS	0.23	26.17	11960	< .001	[0.21, 0.25]
Interaction between SR and GM	-0.08	-9.57	11940	< .001	[-0.10, -0.06]

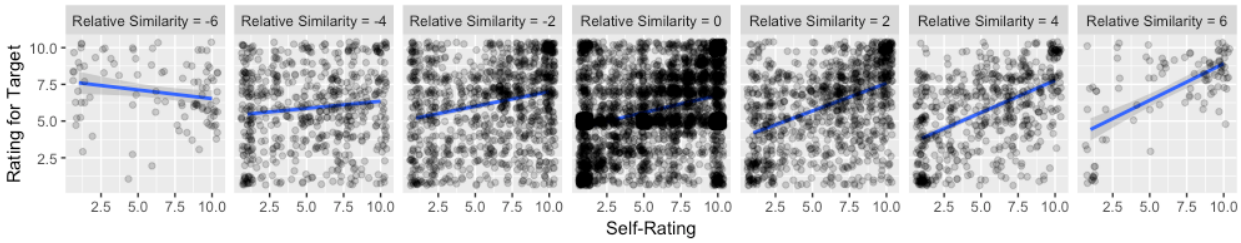
Note. There were significant effects of relative similarity in Study 1 and Study 2 in predicting ratings of a target. That is, the more similar participants perceived themselves to be to the target

versus the other target, the higher the rating participants gave a target on a statement. Importantly, this was qualified by an interaction with self-rating in both studies, and the effect size of the interaction was larger than that of the main effect of relative similarity. Thus, we should interpret this finding in the context of the interaction.

1) Study 1



2) Study 2



3) Study 3

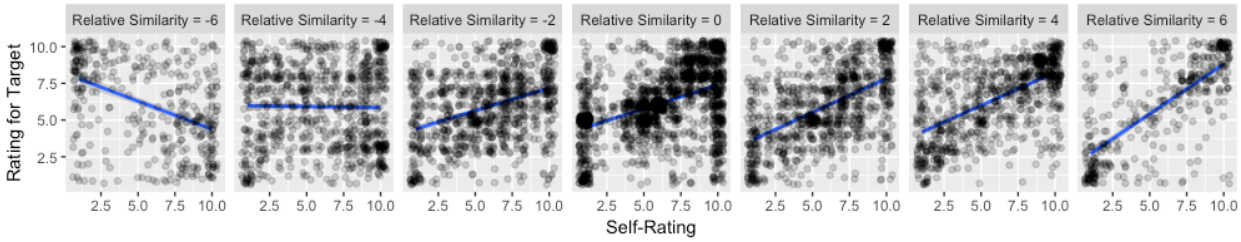
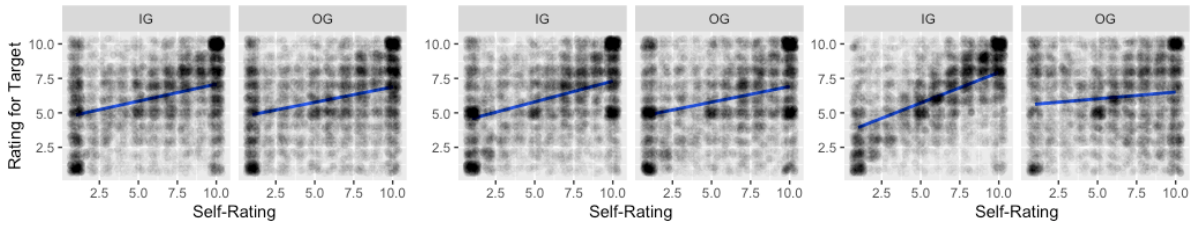


Figure 2. The relationship between self-ratings and ratings for a target at different levels of relative similarity in Studies 1, 2, and 3. Dots represent individual participants' 100 ratings for themselves in relation to their ratings for each target.



(a) (b) (c)

Figure 3. The relationship between self-ratings and ratings for the ingroup and outgroup targets in Studies 1 (a), 2 (b), and 3 (c). Dots represent individual participants' 100 ratings for themselves in relation to their ratings for each target. IG indicates ingroup and OG indicates outgroup.

Incorrectness of Ratings.

Having found variation in simulation as a function of relative similarity, we then conducted planned analyses to examine the extent to which participants were incorrect about others' attitudes in relation to how similar participants felt to a target representative of a group. We created an incorrectness score that represented the absolute difference between a participant's judgment of a target (e.g., the liberal target or the Eagle target) and the mean rating of all participants in that target's group (all participants who identified more strongly with the liberal target or all participants assigned to the Eagles) for each of the 100 statements. (For Study 1, there were an additional 25 participants (*M* age = 35.88 [*SD* = 12.34]; 11 women, 14 men) from pilot data collection whose ratings we included in our calculations of mean ratings of participants in the targets' groups so that they would be more representative ratings of the two groups.)

For each study, we ran mixed-effects models in which the dependent variable was incorrectness; the fixed effects were relative similarity and group membership; and the random

effects were participant ID and item number. Across studies, there were two main effects. First, the more participants perceived themselves to be similar to one target versus the other target, the more correct participants were about the attitudes expressed by actual participants in that target's group (see Table 3 and Figure 4). Second, participants were more correct about attitudes expressed by actual participants in the ingroup target's group than they were about attitudes expressed by actual participants in the outgroup target's group (see Table 3 and Figure 5). Critically, across studies, the effect involving relative similarity was larger in effect size than the effect involving group membership.

Table 3

Relative Similarity and Group Membership as Predictors of Incorrectness

	β	*t*	*df*s	*p*	95% CI
Study 1					
Relative Similarity	-0.13	-16.50	13030	< .001	[-0.15, -0.12]
Group Membership	-0.03	-4.64	13030	< .001	[-0.05, -0.02]
Study 2					
Relative Similarity	-0.05	-6.13	13030	< .001	[-0.06, -0.03]
Group Membership	0.01	2.29	13030	.021	[0.002, 0.03]
Study 3					
Relative Similarity	-0.13	-14.59	12040	< .001	[-0.15, -0.11]
Group Membership	0.02	2.77	12040	.005	[0.007, 0.04]

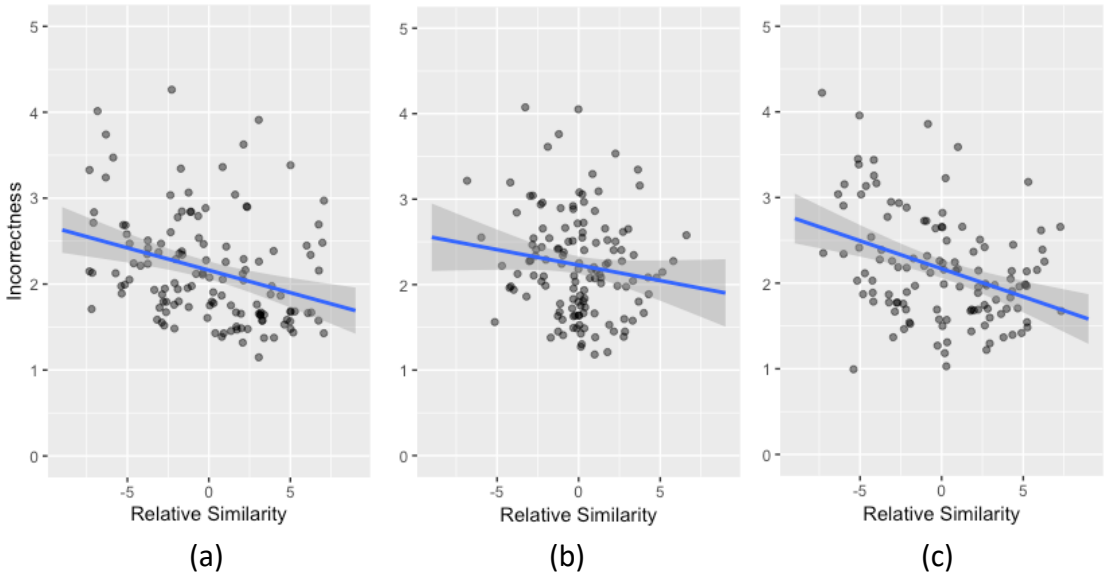


Figure 4. The relationship between relative similarity and incorrectness in Studies 1 (a), 2 (b), and 3 (c). Dots represent the mean incorrectness of each participant for the two targets.

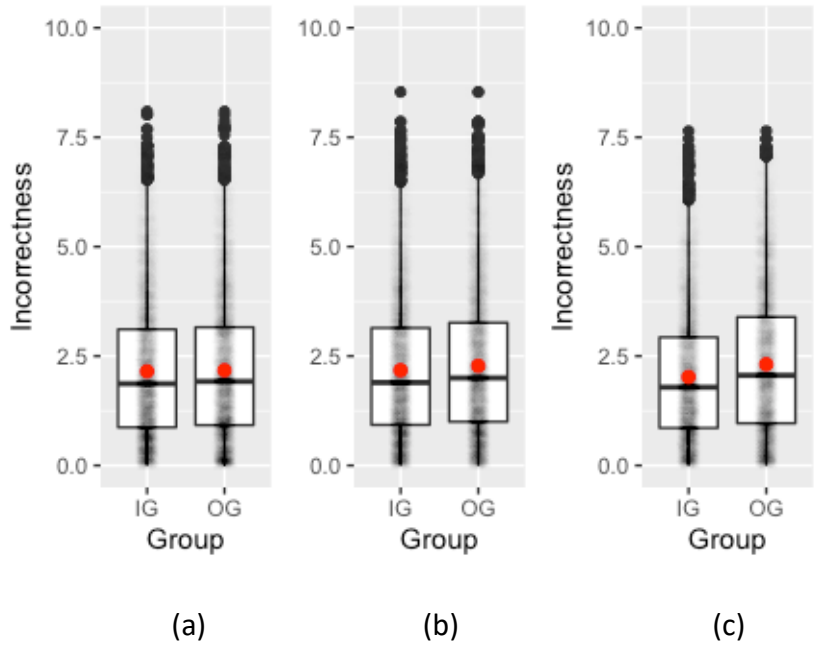


Figure 5. The relationship between group membership and incorrectness in Studies 1 (a), 2 (b), and 3 (c). Red dots indicate means. Boxes indicate interquartile ranges and horizontal lines indicate medians. Translucent gray dots represent individual participants' 100 incorrectness scores for each target. IG indicates ingroup and OG indicates outgroup.

Relation Between Atypical Self-Ratings and Ratings for a Target

Although our planned analyses indicated that participants generally simulate more strongly for targets to whom they feel more similar, it may not be reasonable to simulate if perceivers' attitudes are atypical. We conducted exploratory analyses to determine whether perceivers recognize when their attitudes are atypical and simulate less in such circumstances. We first aimed to determine how atypical attitudes actually were, relative to the two groups of actual participants. We created atypicality scores by taking the absolute difference between a participant's rating of an item and the mean rating of all participants in a target's group for each of the 100 statements and for each of the two targets; larger atypicality would indicate that an

attitude is more atypical. Thus, we can determine how atypical participants are in their attitudes relative to each of the two groups of participants.

For each study, we ran a mixed-effects model in which the dependent variable was the rating for a target on a specific item; the fixed effects were the self-rating for the item, relative similarity, atypicality, group membership (ingroup = -0.5, outgroup = 0.5), and all possible interactions (except for interactions between group membership and relative similarity); and the random effects were participant ID and item number. In the models, fixed effects were centered. We focused on two sets of interactions in these models.

First, we examined interactions involving relative similarity. Across studies, the more participants perceived themselves to be relatively more similar to a target versus the other target, the more positively their self-ratings on attitudes predicted their ratings for that target. Additionally, across studies (see Table 4 and Figure 6), this relationship was not moderated by atypicality[1].

Second, we examined interactions involving group membership. In Studies 1 and 3, but not in Study 2, participants' self-ratings on attitudes more strongly predicted their ratings of the ingroup target than those of the outgroup target. This relationship was moderated by atypicality in Study 1, but not in Studies 2 and 3 (see Table 4 and Figure 7). Specifically, when participants' attitudes were low in atypicality, participants' self-ratings were more predictive of

[1] As these analyses were not planned and the present studies were not designed to be powered to assess these questions, these analyses should be interpreted with caution.

their ratings of the ingroup target than those of the outgroup target. By contrast, when participants' attitudes were high in atypicality, this difference in predictiveness was lessened.

Critically, across studies, the interaction between relative similarity and self-ratings was larger in effect size than the interaction between group membership and self-ratings.

Table 4

The Role of Atypicality in Simulation

	β	*t*	*df*s	*p*	95% CI
Study 1					
Interaction between Self-Ratings and RS	0.19	22.84	13030	< .001	[0.18, 0.21]
Interaction between Self-Ratings and GM	0.03	4.08	13040	< .001	[0.01, 0.05]
Three-way Interaction (RS)	0.01	1.12	13040	0.26	[-0.007, 0.02]
Three-way Interaction (GM)	0.01	2.22	13050	0.02	[0.002, 0.03]
Study 2					
Interaction between Self-Ratings and RS	0.14	16.86	13030	< .001	[0.12, 0.15]
Interaction between Self-Ratings and GM	0.002	0.32	13000	.323	[-0.01, 0.01]
Three-way Interaction (RS)	0.015	1.77	13030	.075	[-0.001, 0.03]
Three-way Interaction (GM)	-0.002	-0.32	13030	.744	[-0.01, 0.01]
Study 3					
Interaction between Self-Ratings and RS	0.24	22.68	12030	< .001	[0.21, 0.26]
Interaction between Self-Ratings and GM	-0.07	-7.00	12030	< .001	[-0.09, -0.05]
Three-way Interaction (RS)	-0.01	-1.49	12030	.134	[-0.03, 0.004]
Three-way Interaction (GM)	-0.01	-1.43	12000	.153	[-0.03, 0.003]

Note. RS indicates relative similarity and GM indicates group membership. Three-way interaction indicates the interaction between self-ratings, atypicality, and either RS or GM.

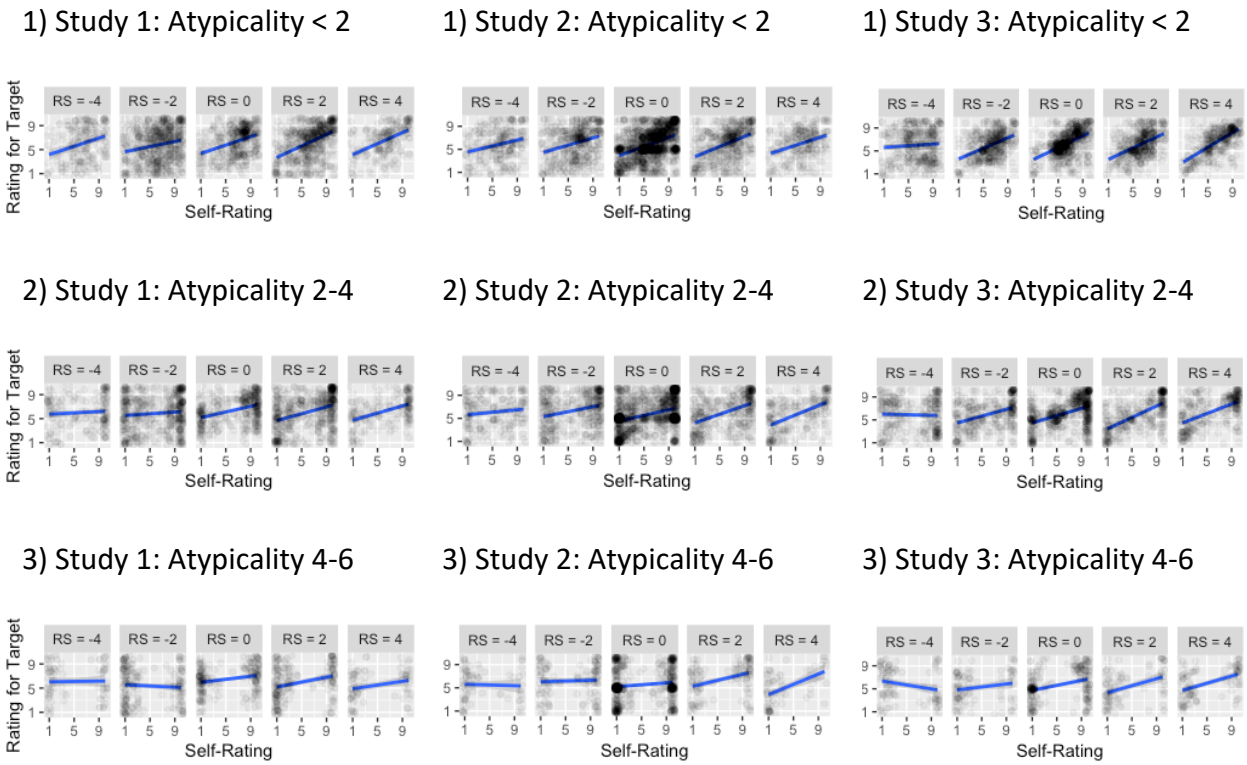


Figure 6. The relationship between self-ratings and ratings for a target at different levels of relative similarity and at different degrees of atypicality in Studies 1, 2, and 3. Dots represent individual participants' 100 ratings for themselves in relation to their ratings for each target. RS indicates relative similarity.

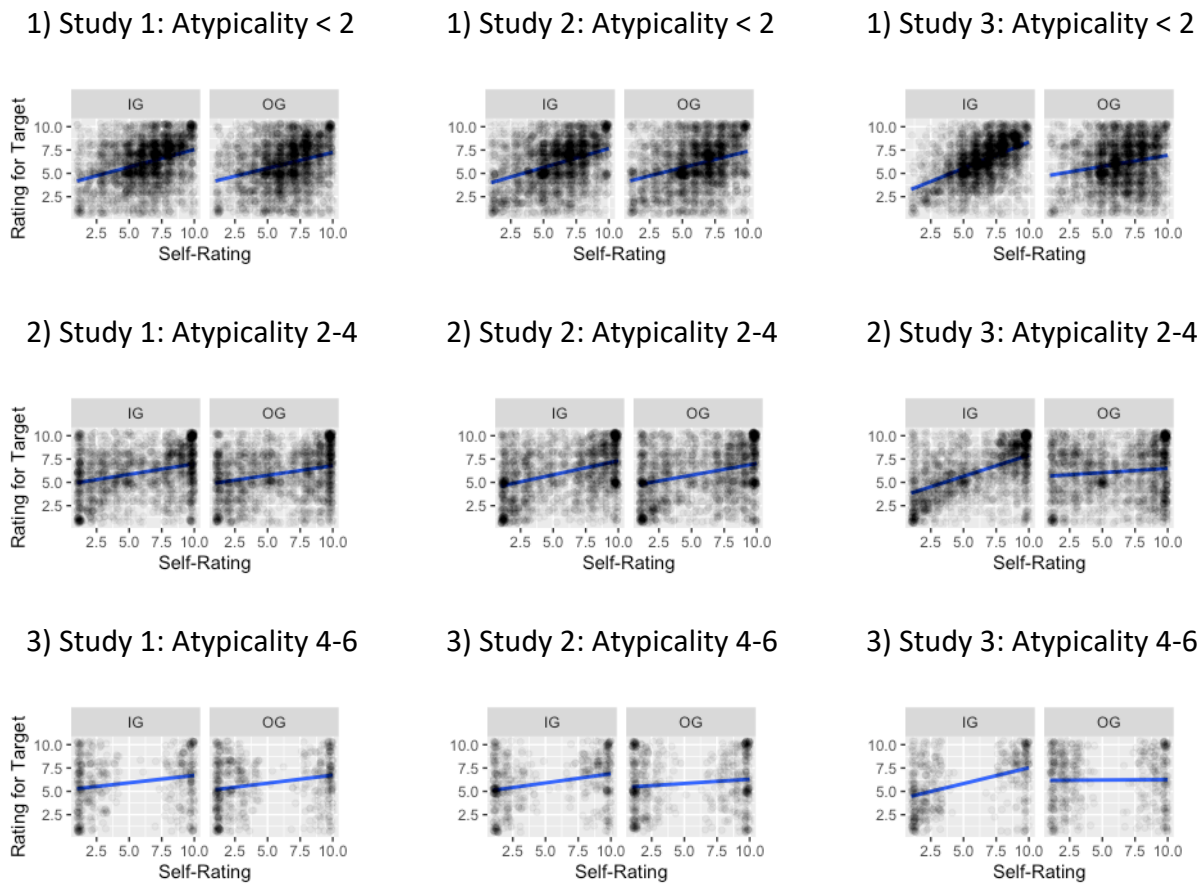


Figure 7. The relationship between self-ratings and ratings for a target at different levels of relative similarity and in relation to group membership in Studies 1, 2, and 3. Dots represent individual participants' 100 ratings for themselves in relation to their ratings for each target. RS indicates relative similarity, IG indicates ingroup, and OG indicates outgroup.

Incorrectness of Ratings by Relative Similarity and Atypicality

In our planned analyses, we found that participants are more correct about the attitudes of relatively more similar others, the individuals that they more strongly simulate. The present analysis examined whether participants' correctness about the attitudes of relatively more similar others may differ when participants' attitudes are atypical.

For each study, we ran mixed-effects models in which the dependent variable was incorrectness; the fixed effects were relative similarity, group membership (ingroup = -0.5, outgroup = 0.5), atypicality, and all possible interactions (except those involving both group membership and relative similarity); and the random effects were participant ID and item number. We focused on two sets of interactions in these models.

First, we examined interactions involving relative similarity. Across studies, the more participants perceived themselves to be relatively more similar to a target versus the other target, the more correct participants were about the attitudes expressed by actual participants in that target's group (see Table 4 and Figure 8). Notably, this main effect of relative similarity was qualified by an interaction with atypicality: Across studies, the more participants perceived themselves to be relatively more similar to a target versus the other target, the more positively atypicality predicted incorrectness.

Second, we examined interactions involving group membership. Across studies, participants were more correct about the attitudes expressed by actual participants in the ingroup target's group than they were about the attitudes expressed by actual participants in the outgroup target's group (see Table 4 and Figure 9). This main effect was qualified by an interaction in Study 3, but not Studies 1 and 2. Specifically, in Study 3, atypicality predicted incorrectness more positively for attitudes expressed by actual participants in the ingroup target's group than for attitudes expressed by actual participants in the outgroup target's group.

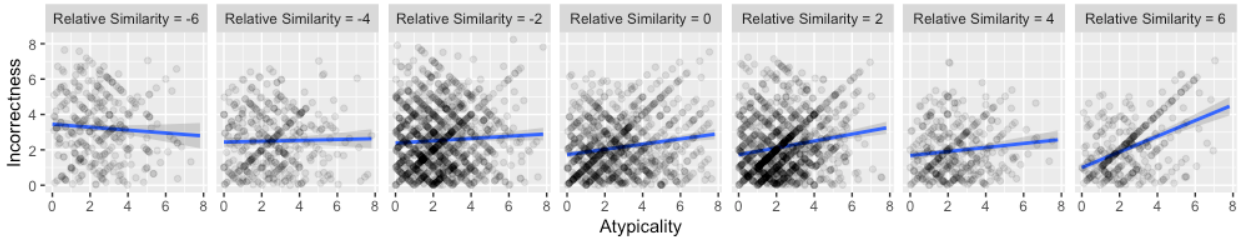
Critically, across studies, the interaction between perceived similarity and atypicality was larger in effect size than the interaction between group membership and atypicality.

Table 4
The Role of Atypicality in Predicting Incorrectness

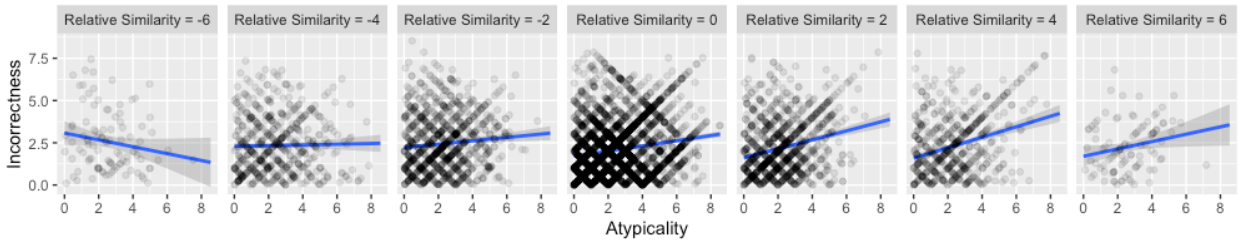
	β	*t*	*df*s	*p*	95% CI
Study 1					
Atypicality	0.13	15.83	13160	< .001	[0.11, 0.15]
Relative Similarity	-0.14	-16.82	13030	< .001	[-0.15, -0.12]
Group Membership	-0.04	-5.42	13040	< .001	[-0.06, -0.02]
Interaction between Atypicality and RS	0.09	11.61	13040	< .001	[0.08, 0.11]
Interaction between Atypicality and GM	0.01	1.36	13040	.171	[-0.004, 0.02]
Study 2					
Atypicality	0.13	15.54	13090	< .001	[0.11, 0.15]
Relative Similarity	-0.04	-6.03	13030	< .001	[-0.06, -0.03]
Group Membership	0.02	2.46	13040	.013	[0.004, 0.03]
Interaction between Atypicality and RS	0.07	9.50	13040	< .001	[0.06, 0.09]
Interaction between Atypicality and GM	-0.004	-0.51	13040	.606	[-0.02, 0.01]
Study 3					
Atypicality	0.18	20.52	12120	< .001	[0.16, 0.19]
Relative Similarity	-0.14	-15.39	12000	< .001	[-0.16, -0.12]
Group Membership	0.01	2.06	12040	.039	[0.0009, 0.03]
Interaction between Atypicality and RS	0.09	10.79	12040	< .001	[0.08, 0.11]
Interaction between Atypicality and GM	-0.04	-5.08	12040	< .001	[-0.06, -0.02]

Note. RS indicates relative similarity and GM indicates group membership.

1) Study 1



2) Study 2



3) Study 3

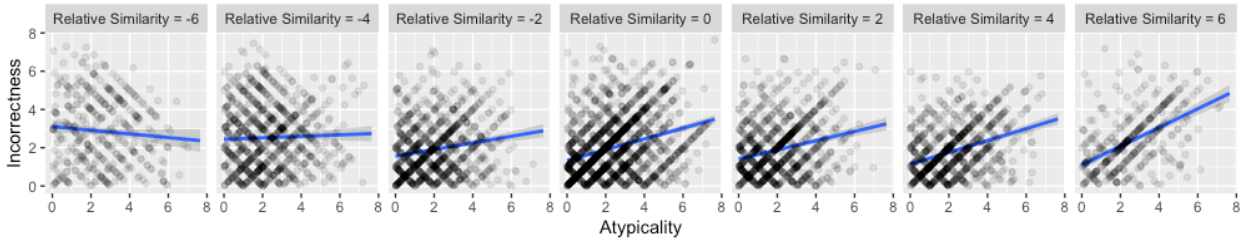
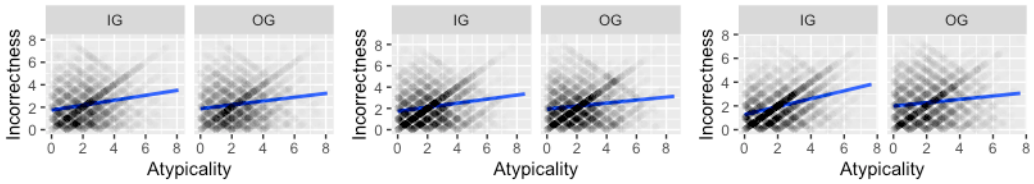


Figure 8. The relationship between atypicality and incorrectness at different levels of relative similarity in Studies 1, 2, and 3. Dots represent individual participants' 200 atypicality scores in relation to their incorrectness for each target.



(a) (b) (c)

Figure 9. The relationship between atypicality and incorrectness in relation to group membership in Studies 1 (a), 2 (b), and 3 (c). Dots represent individual participants' 200 atypicality scores in relation to their incorrectness for each target. IG indicates ingroup and OG indicates outgroup.

Discussion

Across three studies using different group manipulations, we examined whether group membership and the extent to which participants felt similar to a target representative of a group predicted the relation between participants' own attitudes and their judgments of others' attitudes. Across all studies, despite groups differing in attitudes at about chance levels and despite groups not being associated with higher relative similarity in Study 2, the more strongly participants felt similar to a target, the more strongly participants' own attitudes predicted their ratings of the target's attitudes. That is, we replicated research on stronger simulation for more similar others (Tamir & Mitchell, 2013), even though these attitudes appeared irrelevant to the nature of the groups that provided a basis for similarity. Additionally, we extended past work by providing evidence that participants simulate when reasoning about a person's attitudes more strongly and reliably because they perceive themselves to be similar to the person than because they are in the same group as the person. This is consistent with the idea that shared group membership may influence simulation insofar as it causes people to feel more similar to others.

This replication and extension of past work in multiple studies enabled us to address our central question of whether perceivers are any more correct about another's attitudes for more similar others, the individuals whom they more strongly simulate. Across all studies, the more participants felt similar to a target representative of a group, the more correct participants were; that is, the more that participants' ratings of a target's attitudes approximated those of the actual group. Although participants were also more correct when the target was in participants' own group than when the target was in the participants' outgroup, the effect of perceived similarity was larger than that of group membership. The answer to our question, then, was yes: Perceivers are more correct about another's attitudes for more similar others—the individuals whom they more strongly simulate.

The present findings are compelling in multiple ways. First, in using different group manipulations across the three studies, we demonstrate that our findings generalize. Second, in establishing groups based on whether birthdates were odd- or even-numbered in Study 2 and on random assignment (ostensibly based on a personality test) in Study 3, we demonstrate that effects were not isolated to real political groups. In Study 1, participants may have been more correct about the attitudes of groups with representatives to whom they felt relatively more similar because they likely have had even more experiences with people of a similar political orientation. This difference in experiences with groups could not explain the pattern of findings in Study 2 as strongly, given that people do not as strongly publicly identify as having an odd- or even-numbered birthdate, and could certainly not explain the pattern of findings in Study 3, given that groups were randomly assigned there (i.e., no prior experience with the groups would be meaningful).

Past work has presented simulation as a method through which perceivers can reason about the mental states of more similar others (Ames, 2004a, 2004b; Mitchell et al., 2005). Overall, the present studies offer a thorough characterization of the simulation of similar but not dissimilar others, and the limitations of such selective simulation for mental state understanding. Taken together, the present findings are suggestive that simulation as a method of understanding more similar others' attitudes occurs readily, but that it may not be appropriate to avoid simulation for dissimilar others. Although perceivers do not use their own attitudes to predict those of targets that they feel dissimilar to, they may be wrong to not do so. Future work may examine the implications of this erroneous avoidance of simulation for more dissimilar others. Past work has demonstrated that people like others more if they share more attitudes (Byrne, 1969, 1971; Byrne & Nelson, 1965; Kaplan & Anderson, 1973; Rosenbaum, 1986; Singh & Ho, 2000). If people falsely assume that outgroup members are unlike the self in attitudes, even when attitudes are irrelevant to the group, there may be less cooperation between people of different groups. People of different social groups may be less integrated with each other, despite their attitudes not always differing strongly.

How might perceivers manage to more accurately infer the opinions and attitudes of similar versus dissimilar others? Critically, the attitudes in our study did not differ by group in any of the three studies. This is suggestive that the difference between different groups (liberals versus conservatives; Eagles versus Rattlers) in attitudes may not have been as large as perceivers might intuitively think. It may be more appropriate to simulate, regardless of how different a perceiver thinks that similar others are from dissimilar others, then, for attitudes that are irrelevant to the group that may form the basis of perceiving similarity (e.g., how much liberals versus conservatives enjoy exercise).

This is not to say that there are never differences between groups in attitudes, only that there were none in the present studies, and that there may often be none in real life when attitudes are irrelevant to group membership. This was largely by design in the present studies, as we were focused on intrusive simulation: simulation where the group information is uninformative. Past work has examined simulation of attitudes that are relevant to the group (e.g., Chambers et al., 2006). The goal of the present studies was to focus on attitudes that are irrelevant to the group. These may not be attitudes that are associated with the group, but attitudes that people could still reason about in real life, and potentially incorrectly. Although these attitudes did not discriminate between groups in the present studies, they could potentially in scenarios in which the groups actually differ in attitudes. That is, in the present studies, participants may have been in different groups (Liberals versus Conservatives), but still had a lot of things in common, like being based in the United States; a participant's attitude towards something like ice cream, then, might be easily generalizable to others, regardless of which group someone belongs to. By contrast, there could be cases in which such simulation is inappropriate (e.g., to a culture in which people are largely lactose intolerant or do not eat milk-products). Future work should examine whether not simulating for more dissimilar others may be correct in scenarios in which there is more reason to think that one's group and another's group actually differ in attitudes.

This is also not to say, of course, that any given perceiver will not have atypical or extreme views on a topic, only that across such topics, most perceivers will have opinions that are average. In other words, if perceivers know nothing about the distribution of opinions in the world and are not atypical in their views, then they may be better off assuming that other individuals share their opinions on any particular topic. If perceivers are atypical, though, then it

may be erroneous to simulate for others. For example, if a perceiver despises ice cream while most other people in a perceiver's social group love it, then it would be incorrect for the perceiver to simulate and assume that all others are like the self, despite group membership being shared. Our exploratory analyses suggest that when perceivers' attitudes are atypical, perceivers are especially wrong about the attitudes of more similar others; perceivers are not decreasing in simulation when their attitudes are atypical. In one of our three studies, we find evidence that participants simulate less strongly when their attitudes are more atypical for ingroup members; however, this effect is inconsistent and was not replicated across studies. Further, we find that in only one of the three studies that perceivers are more wrong about ingroup attitudes than outgroup attitudes when their attitudes are more atypical. This difference in the effects of perceived similarity and group membership is consistent with the idea that perceived similarity influences simulation more strongly than does group membership.

A general tendency to simulate for more similar others even when attitudes are atypical could lead to tension between the self and more similar others, if perceivers falsely assume that similar others are like them in atypical attitudes. In addition to seeking to replicate these findings on atypical attitudes, future work should examine if perceivers simulate less strongly for more similar others when they recognize that their attitudes are atypical. If perceivers do not recognize when their attitudes are atypical and fail to adjust accordingly, future work may examine interventions that may correct this erroneous simulation of atypical attitudes.

In sum, the present studies both conceptually replicate and build on past work on simulation. The present studies demonstrate that the more that people feel similar to a target's group, the closer people's assumptions of a target's attitudes are to the actual attitudes of participants in the target's group. That is, people are more correct about the attitudes of targets

to whom they feel more similar—the targets that they more strongly simulate. Additionally, the present studies qualify this, by demonstrating that simulation for more similar others occurs even when people's attitudes are atypical. These findings demonstrate limitations to the strategy of selectively simulating for more similar others.

Open Practices

Data, stimuli, materials, and analysis scripts are available on the Open Science Framework (OSF) at https://osf.io/pmkh6/?view_only=6f56e61caa834c0c9cb32ac366eafc52.

References

Ames, D. R. (2004a). Inside the mind reader's tool kit: Projection and stereotyping in mental state inference. *Journal of Personality and Social Psychology*, 87, 340-353.

Ames, D. R. (2004b). Strategies for social inference: A similarity contingency model of projection and stereotyping in attribute prevalence estimates. *Journal of Personality and Social Psychology*, 87, 573-585.

Ames, D. R., Mor, S., & Toma, C. (2013). The double-edge of similarity and difference mindsets: What comparison mindsets do depends on whether self or group representations are focal. *Journal of Experimental Social Psychology*, *49*(3), 583-587.

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face images. *Journal of Experimental Psychology: General, 142*(4), 1323-1334.

Bruneau, E. G., Cikara, M., & Saxe, R. (2015). Minding the gap: Narrative descriptions about mental states attenuate parochial empathy. *PLoS ONE*, *10*(10), e0140838.

Byrne, D. (1969). Attitudes and attraction. In L. Berkowitz (Ed.), *Advances in experimental socialpsychology* (Vol. 4, pp. 35-89). New York: Academic Press.

Byrne, D. (1971). *The attraction paradigm*. New York: Academic Press.

Byrne, D., & Nelson, D. (1965). Attraction as a linear function of proportion of positivereinforcements. *Journal of Personality and Social Psychology*, 1, 659-663.

Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy*, *97*(5), 1232-1254.

Clement, R. W., & Krueger, J. (2002). Social categorization moderates social projection. *Journal of Experimental Social Psychology*, *38*(3), 219-231.

Chambers, J. R., Baron, R. S., & Inman, M. L. (2006). Misperceptions in intergroup conflict: Disagreeing about what we disagree about. *Psychological Science*, *17*(1), 38-45.

Cho, J. C., & Knowles, E. D. (2013). I'm like you and you're like me: Social projection and self-stereotyping both help explain self–other correspondence. *Journal of personality and social psychology*, *104*(3), 444-456.

Cronbach, L. J. (1955). Processes affecting scores on" understanding of others" and" assumed similarity.". *Psychological bulletin*, *52*(3), 177-183.

Davies, M., & Stone, T. (1995a). *Folk Psychology: The Theory of Mind Debate*. Oxford, UK: Blackwell Publishers.

Davies, M., & Stone, T. (1995b) *Mental Simulation: Evaluations and Applications*, Oxford, UK: Blackwell Publishers.

Dunham, Y. (2018). Mere membership. *Trends in Cognitive Sciences*, *22*(9), 780-793.

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

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*(12), 493-501.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological review*, *109*(1), 3-25.

Harris, L. T., & Fiske, S. T. (2006). Dehumanizing the lowest of the Low: Neuroimaging Responses to Extreme Out-Groups. *Psychological Science*, 17(10), 847–853.

Heal, J. (1986). Replication and functionalism. In Butterfield, J. (Ed.), *Language, Mind and Logic* (pp. 135-150).

Hoch, S. J. (1987). Perceived consensus and predictive accuracy: The pros and cons of projection. *Journal of Personality and Social Psychology*, *53*, 221–234.

Ichheiser, G. (1949). Misunderstandings in human relations: A study in false social perception. *American Journal of Sociology*, 70.

Jenkins, A. C., Macrae, C. N., & Mitchell, J. P. (2008). Repetition suppression of ventromedial prefrontal activity during judgments of self and others. *Proceedings of the National Academy of Sciences*, *105*(11), 4507-4512.

Kaplan, M. F., & Anderson, N. H. (1973). Information integration theory and reinforcement theory asapproaches to interpersonal attraction. Journal of Personality and Social Psychology, 28 , 301-312.

Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: a social relations analysis. *Psychological bulletin*, *102*(3), 390-402.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25-41.

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, *104*(30), 12577-12580.

Liberman, Z., Woodward, A. L., & Kinzler, K. D. (2017). Preverbal infants infer third-party social relationships based on language. *Cognitive Science*, *41*, 622-634.

Meltzoff, A. N., and Brooks, R. (2001). ''Like me'' as a building block for understanding other minds: Bodily acts, attention, and intention. In Malle, B. F., Moses, L. J., and Baldwin, D. A. (Eds.), *Intentions and Intentionality: Foundations of Social Cognition* (pp. 171-191). Cambridge, MA: MIT Press.

Mitchell, J. P., Banaji, M. R., & MacRae, C. N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience*, *17*(8), 1306-1315.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, *50*(4), 655-663.

Mor, S., Toma, C., Schweinsberg, M., & Ames, D. (2019). Pathways to intercultural accuracy: Social projection processes and core cultural values. *European Journal of Social Psychology*, *49*(1), 47-62.

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, *125*, 737-759.

O'Brien, E., & Ellsworth, P. C. (2012). More than skin deep: Visceral states are not projected onto dissimilar others. *Psychological Science*, *23*(4), 391-396.

Otten, S. (2016). The Minimal Group Paradigm and its maximal impact in research on social categorization. *Current Opinion in Psychology*, *11*, 85-89.

Robbins, J. M., & Krueger, J. I. (2005). Social projection to ingroups and outgroups: A review and meta-analysis. *Personality and Social Psychology Review*, *9*(1), 32-47.

Rosenbaum, M. E. (1986). The repulsion hypothesis: On the nondevelopment of relationship. *Journal of Personality and Social Psychology*, 51, 1156-1166.

Saxe, R. (2005). Against simulation: the argument from error. *Trends in Cognitive Sciences*, *9*(4), 174-179.

Singh, R., & Ho, S. Y. (2000). Attitudes and attraction: A new test of the attraction, repulsion and similarity-dissimilarity asymmetry hypotheses. *British Journal of Social Psychology*, *39*(2), 197-211.

Stone, T., & Davies, M. (1995). *Mental Simulation: Evaluations and Applications*. Blackwell.

Swann, W. B. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review*, *91*(4), 457-477.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, *223*(5), 96-103.

Tamir, D. I., & Mitchell, J. P. (2013). Anchoring and adjustment during social inferences. *Journal of Experimental Psychology: General*, *142*(1), 151-162.

Todd, A. R., Simpson, A. J., & Tamir, D. I. (2016). Active perspective taking induces flexible use of self-knowledge during social inference. *Journal of Experimental Psychology: General*, *145*(12), 1583-1588.

Turner, J. C. (1985). Social categorization and the self-concept: a social cognitive theory of group behavior. In E. J. Lawler (Ed.), *Advances in Group Processes: Theory and Research* (pp. 77-121). Greenwich, CT: JAI Press.

Turner, J. C., & Oakes, P. J. (1986). The significance of the social identity concept for social psychology with reference to individualism, interactionism and social influence. *British Journal of Social Psychology*, *25*(3), 237-252.

Turner, J.C. and Oakes, P.J. (1989) Self-categorization theory and social influence. In P.B. Paulus (Ed.), *The Psychology of Group Influence* (pp. 233-275). Hillsdale, NJ: Erlbaum.

Van Boven, L., Dunning, D., & Loewenstein, G. (2000). Egocentric empathy gaps between owners and buyers: misperceptions of the endowment effect. *Journal of Personality and Social Psychology*, *79*(1), 66-76.

Van Veelen, R., Otten, S., Cadinu, M., & Hansen, N. (2016). An integrative model of social identification: Self-stereotyping and self-anchoring as two cognitive pathways. *Personality and Social Psychology Review*, *20*(1), 3-26.

Xu, X., Zuo, X., Wang, X., & Han, S. (2009). Do you feel my pain? Racial group membership modulates empathic neural responses. *Journal of Neuroscience*, *29*(26), 8525-8529.