American Psychological Association

Journal of Experimental Psychology: General

0096-3445/18/$12.00

http://dx.doi.org/10.1037/xge0000470

Discovering Social Groups via Latent Structure Learning

Tatiana Lau, Hillard T. Pouncy, Samuel J. Gershman, and Mina Cikara
Harvard University

Humans form social coalitions in every society on earth, yet we know very little about how social group boundaries are learned and represented. We derive predictions from a computational model of latent structure learning to move beyond explicit category labels and mere similarity as the sole inputs to social group representations. Four experiments examine (a) how evidence for group boundaries is accumulated in a consequential social context (i.e., learning about others' political values), (b) to what extent learning about these boundaries drives one's own choices as well as attributions about other agents in the environment, and (c) whether these latent groups affect choice even in the presence of group labels that contradict the latent group structure. Our results suggest that people integrate information about how agents in the environment relate to *one another* in addition to oneself to infer social group structure. We argue that this mechanism is a plausible explanation of other theories of social relations—for example, balance theory.

Keywords: latent structure learning, social categorization, social inference

Supplemental materials: http://dx.doi.org/10.1037/xge0000470.supp

Social groups are a universal feature of human societies and drive a vast array of decisions: from the most quotidian—who do I approach at this party?—to the most consequential—whom do we target with missile strikes? In many cases, people utilize category labels (e.g., gender, nation of origin, arbitrary assignment to novel groups) and generic language or visual cues (e.g., skin tone) to determine how to evaluate others, how to allocate their resources, and what norms to follow in social settings (e.g., Allport, 1954; Brewer, 1999; Gelman, Collman, & Maccoby, 1986; Hewstone, Rubin, & Willis, 2002; Kubota, Banaji, & Phelps, 2012; Rhodes, Leslie, & Tworek, 2012; Tajfel et al., 1971; Taylor & Gelman, 1993; Terry & Hogg, 1996). In the absence of overt cues, both adults and young children frequently use judgments of familiarity or similarity to one's self on some feature (e.g., attitudes, accents) to guide their social preferences (e.g., Byrne & Nelson, 1965; Kinzler, Shutts, Dejesus, & Spelke, 2009; Rokeach, Smith, & Evans, 1960).

Classic theories of intergroup relations highlight several other dimensions by which groups are defined: namely, common fate within groups (Campbell, 1958) and functional relations between groups (Sherif, 1966). In other words, people have strong expectations about the nature of obligations within and between groups (e.g., Brewer, 1999, 2008; Rhodes & Chalik, 2013), such that ingroup members' outcomes are perceived to be interdependent (e.g., "a policy change that affects me will affect my entire group") and often at odds with the outgroup (e.g., "what is good for us is bad for them"). Thus, another process through which people may categorize others as ingroup or outgroup members is by inferring how agents in the environment relate to one another in addition to oneself (Heider, 1958; see also Chang, Krosch, & Cikara, 2016; Cikara & Van Bavel, 2014; Cikara, Van Bavel, Ingbretsen, & Lau, 2017; Fiske & Ruscher, 1993). This phenomenon is starkly reflected in coalitions organized around political values and preferences, particularly as specific policy preferences become less well delineated by party lines. These groupings matter; recent evidence suggests that implicit bias and behavioral discrimination along political preferences is now as potent as bias against racial outgroups in some domains (Iyengar, Sood, & Lelkes, 2012; Iyengar & Westwood, 2015; Motyl, Iyer, Oishi, Trawalter, & Nosek, 2014).

How do people accumulate coalition or group structure information from their environments in the absence of overt cues to group membership or in the presence of ambiguous cues? We have recently proposed a formal account of social influence (i.e., the effects of others' preferences on one's own preferences) in multiagent settings. Specifically, we argue that social influence can be better accounted for by a structure learning framework than dyadic similarity-based explanations (Gershman, Pouncy, & Gweon, 2017). Rather than imitating individual agents who seem similar to themselves, people make inferences about latent groupings of others and their corresponding preferences. Bayes' rule provides a normative mechanism for combining observed preferences with prior beliefs to infer a posterior distribution over possible latent groupings (Figure 1).

Tatiana Lau, Hillard T. Pouncy, Samuel J. Gershman, and Mina Cikara, Department of Psychology, Harvard University.

This research was supported by a Mind, Brain, and Behavior Faculty Research Grant awarded to Mina Cikara and Samuel J. Gershman. Tatiana Lau presented these results at the annual meetings for the Society for Neuroeconomics (2017), Society for Judgement and Decision Making (2017), and Society for Personality and Social Psychology (2018). The authors thank Zach Ingbretsen for his assistance programming these experiments.

Correspondence concerning this article should be addressed to Mina Cikara, Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138. E-mail: mcikara@fas.harvard.edu

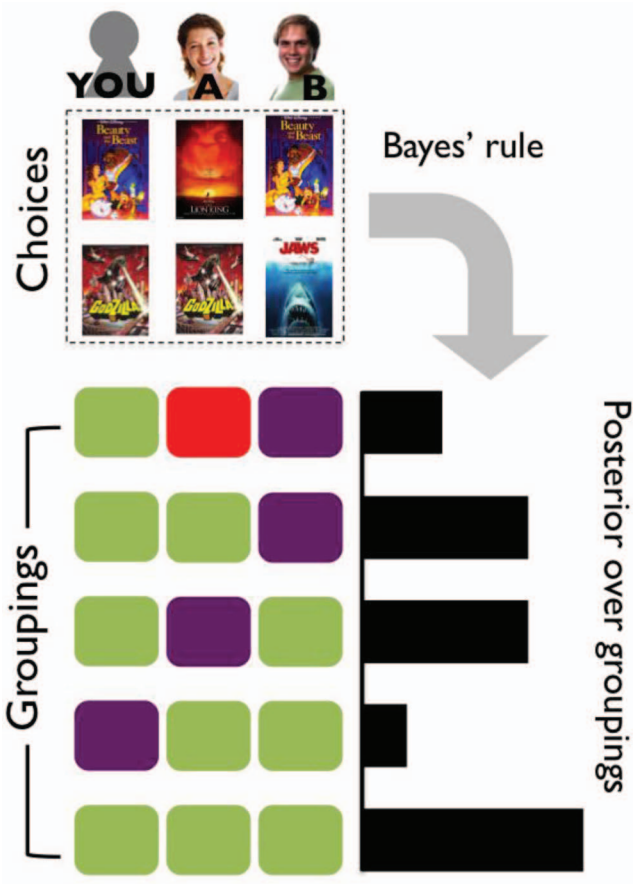

Figure 1. Model schematic illustrating how choice patterns are transformed using Bayes' rule to form a posterior over different possible latent groupings of agents. See the online article for the color version of this figure.

In previous work (Gershman et al., 2017), we tested this hypothesis by asking participants to report their preferences for different movies and to learn the preferences of unfamiliar agents. In a final "mystery choice" trial, participants were given the choice between two unlabeled movies, one favored by Agent A and one favored by Agent B, both of whom agreed with the participant's own prior choices equally often. If people only use dyadic similarities to predict their own preferences, then they should be indifferent between the two movies. However, if they build representations of latent group structure, then their decisions should be influenced by the presence of a third agent (C) during the training phase. If C's choices agree with both the participant and with B more often than with A (Figure 2, left), then the structure learning model predicts that B will exert greater influence on the participant's mystery choice relative to when C's choices agree often with B but not the participant (Figure 2, right). This prediction was confirmed across a series of experiments.

Overview and Hypotheses

The current experiments investigate whether structure learning provides a plausible cognitive mechanism by which people build representations of social groups. Our approach seeks to move beyond explicit category labels and mere similarity as the sole inputs to social group representations, appealing instead to a domain-general structure learning mechanism (Austerweil, Gershman, Tenenbaum, & Griffiths, 2015) that we believe also supports social categorization and intergroup bias. We have chosen to test the limits of this structure learning framework using political stimuli. While movie preferences may not bear much consequence (given that movie preferences typically only affect oneself), political policy preferences reveal one's own values and opinions as to how society should function. Moreover, given today's political climate, participants may be more willing to ascribe greater weight to specific policy preferences that they can more easily identify as one of their own (e.g., Lelkes & Sniderman, 2016), which could, in turn, eliminate the social learning effects found in earlier studies.

Here we test whether structure learning influences choices when the groups are based on political values (Experiments 1 and 2), influences trait judgments of agents (Experiment 3), and induces effects that persist in the presence of explicit, countervailing group labels (Experiment 4). We predict that across all experiments, participants will align themselves with Agent B when Agent C creates a latent group including the participant and B, even though Agents A and B are equally similar to the participant.

Note, however, that the structure learning model from which we derive our hypotheses is impartial between disagreement and agreement. That is, agreement with one agent increases the probability of being clustered with that agent, but disagreement will likewise decrease the probability of clustering with that agent. Therefore, the model also predicts a preference for A (50.05%) over B (49.95%) after blocks in which C's preferences do not align with the participant's. In other words, after blocks of high disagreement between C and the participant, A—the agent that does not belong in a group with B and C—should exert a greater influence on the participant's mystery choice. This alignment mirrors contexts in which social group boundaries are highlighted via disidentification—self-categorization into one group largely in opposition to the alternative group (e.g., Zhong, Phillips, Leonardelli, & Galinsky, 2008). For example, one may not be in love with the Democratic Party, but one identifies with them because Republicans are worse. This disidentification can, in turn, lead to individuals taking political action against the group (e.g., the National Rifle Association) with which they've disidentified (Elsbach & Bhattacharya, 2001). This method has also historically been used to shift political allegiances within minority communities in the United States (Patton, 1995): for example, framing

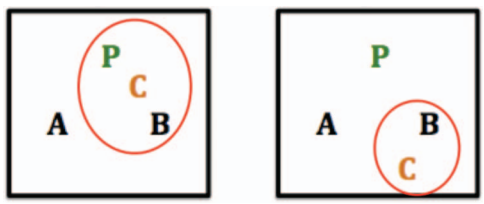

Figure 2. Agents are represented as letters in an abstract space (P = participant), where the distance between letters indicates the degree to which agents agree in their choices (i.e., choice overlap). Red circles indicate the latent groups that have high posterior probability. See the online article for the color version of this figure.

Asian and Latino voters' identity as "not Caucasian" shifted their preferences away from Hillary Clinton to Barack Obama during the 2008 Democratic Party primary (Zhong, Galinsky, & Unzueta, 2008). Similarly, the structure learning model predicts that such an effect may occur after blocks in which Agent C disagrees with many of the participant's preferences.

Complete materials, data, and data analysis code for all four experiments are available for download at OSF: https://osf.io/ay8kg/

Experiment 1: Latent Group Structure Based on Political Issue Preferences Drives Choice

One limitation of previous experiments is that movie preferences have little consequence for others and do not necessarily reveal one's values and opinions or beliefs. Thus one could dismiss our earlier studies using movie preferences as being a category learning study dressed up as social psychology, and therefore not applicable to "real," meaningful social preferences, which cluster around belief and value systems. Here we test whether the latent group structure effects extend to a domain in which people are likely to attend closely to others' preferences as an indicator of their political party group membership—a likely boundary condition on the generalizability of nonsocial category learning.

Method

Participants and exclusions. Gershman et al. (2017) recruited an average of 138 participants per experiment (sufficient sample size to detect a small effect in a within-subjects design). Thus, we recruited 166 participants via Amazon Mechanical Turk (AMT). We excluded five participants for not identifying as male or female (1 person) or for stating that they thought the entire experiment was fake (4 people). We excluded an additional five participants who failed all four questions of the political questionnaire (see the Procedure section), since this indicated to us that the participant (a) did not care about politics and therefore would not find political preferences informative of an agent's identity, (b) was not aware of political issues to the degree that it would be informative of any agent's identity, or (c) rushed through the task. This left us with a sample size of 156 participants (69 female, M_{age} = 34.51 years, SD = 10.45). We also excluded 33 mystery choice trials (out of 156 participants × 4 blocks = 624 trials) in which participants took longer than 20 s to respond. This left us with 591 valid trials (293 high C-agreement, 298 low C-agreement). Across all experiments, participants provided informed consent; all procedures complied with the university's institutional review board's guidelines.

Materials. To develop stimuli, we used ISideWith.com, a website that helps people determine the political party and/or candidate with which their positions best align based on yes and no responses to nationally relevant, political issues (e.g., "Do you support the death penalty?"). The website also aggregates survey responses and makes this data publicly available (https://isidewith.com/polls). We selected issues that had accumulated at least 500,000 votes and were not strongly skewed to one position (i.e., had approximately 50/50 splits on survey responders' preferences). We focused on these low-consensus issues because agents were going to be randomly assigned to their political preferences in our task, and we wanted to avoid generating completely incoherent preference profiles across a block. We included the 32 issues with the lowest yes-no differences as of October 2016 in the main task (see OSF for complete materials).

On each trial, we displayed the issue as text at the top of the screen. Underneath, we signified a "yes" or "no" response to the issue by superimposing a green check mark or a red "X," respectively, atop an image representing the issue. To avoid confusion, we also displayed the words, "YES" and "NO" underneath the corresponding images. The order of presentation of the 32 issues as well as the sides on which the agreement positions appeared on the screen were randomized for each participant.

In order to eliminate the possibility of any cross-categorization effects, we gender-matched the agents to participants' self-reported gender. For agent pictures, we selected a total of 24 photos from the Chicago Face Database (CFD; Ma, Correll, & Wittenbrink, 2015) and gender-matched agents to the participant. We extracted the pool of "White" faces (based on CFD designations) and eliminated faces based on the norming data provided by the CFD (see below) until 12 female and 12 male faces were left. See the online supplementary materials for details on how we selected the 24 faces.

Procedure. We recruited participants under the pretense of playing a game in which they would tell us about their political issue preferences and learn about others' preferences. After providing demographic information (age, ethnicity, and gender), participants completed a sample trial. Here, they expressed their own opinion on a topic ("Should cartoons include plotlines involving duck-hunting?") by selecting "Yes" or "No" and then guessed and received feedback on the opinions of Bugs and Daffy. After this, participants were guided through a mystery choice trial. We told participants that for these trials, colored boxes (blue and green) with question marks on them would represent two different policy positions. The only information participants had about the boxes were the choices of other agents—the same ones about whose preferences they had just learned. We told participants to select the box they would prefer based on the other agents' choices.

Within the main task, each block consisted of two phases (Figure 3): (a) eight regular trials during which participants expressed their own preferences and learned about others' policy preferences, and (b) a mystery choice trial. Each regular trial began with the participant first seeing a policy position and choosing whether they would support that position. Each participant then learned about the preferences of the other three individuals (heretofore Agents A, B, and C) via feedback. Participants first saw a picture of one of the agents alongside his or her name and were asked to predict the preference of that agent with regard to the policy position (e.g., "Which do you think *Jenny* chose?") and then learned of the agent's choice when a blue arrow pointing to one of the preferences (Yes or No) appeared (Figure 3A). Participants then repeated this prediction and feedback process for the other two individuals. Once this process (self-choice, prediction, and feedback for A; prediction and feedback for B; and prediction and feedback for C) was completed for one policy position, participants started a new regular trial by repeating the process for a new policy position with the same three agents. A table consisting of eight rows and four columns displayed on the right-hand side of the screen recorded the participant's and each agent's responses on each trial. The order of the policy positions and individuals was

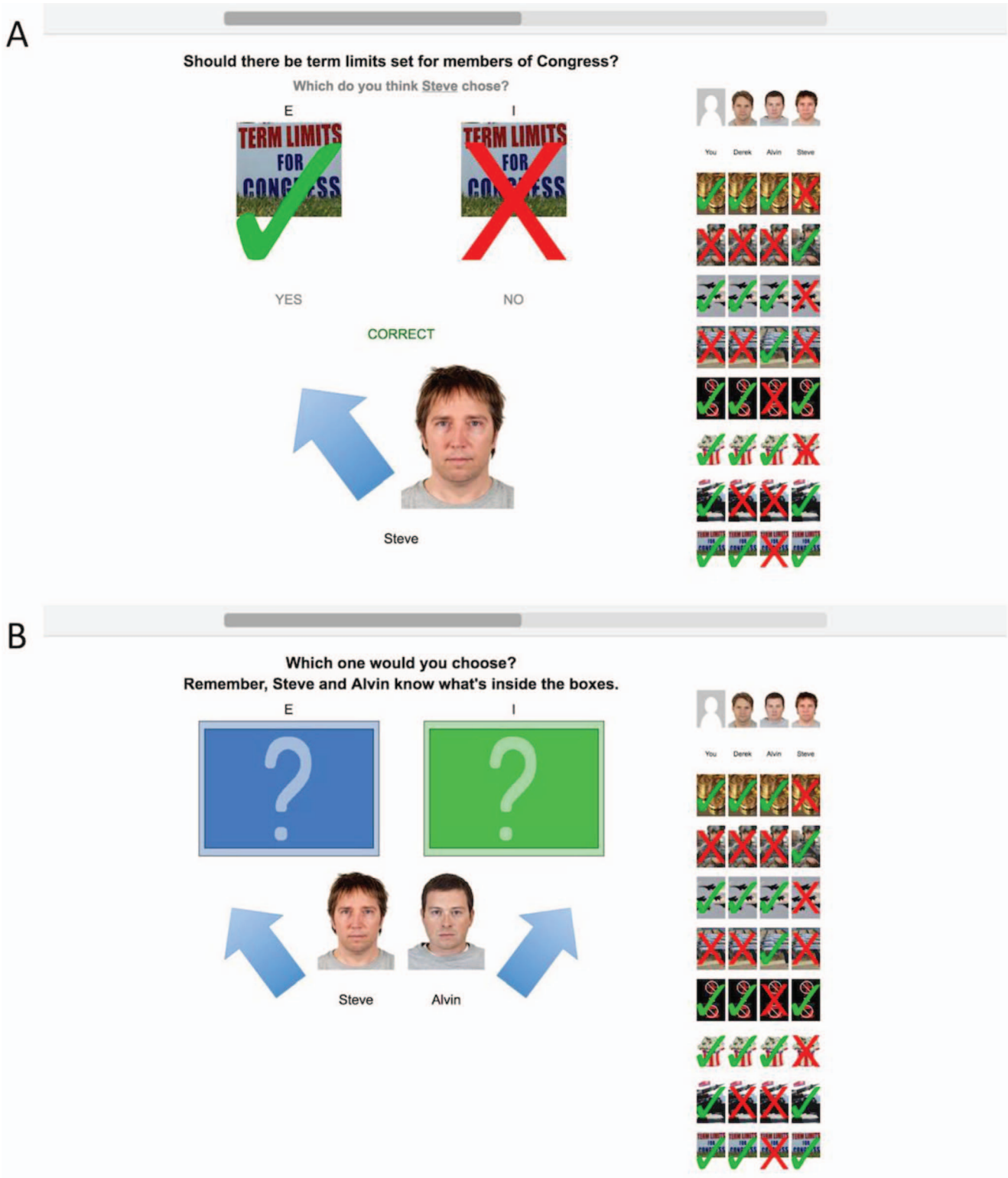

Figure 3. For each trial, participants stated their own political stance and guessed and received feedback on the stances of three agents on that same policy. After doing this eight times (A), participants completed a mystery choice trial, wherein they were asked to choose between Agents A and B (B). Face images from "The Chicago Face Database: A Free Stimulus Set of Faces and Norming Data," By D. S. Ma, J. Correll, & B. Wittenbrink, 2015, *Behavior Research Methods, 47*, pp. 1122–1135. Copyright [2015] by the Psychonomic Society. Reprinted with permission. See the online article for the color version of this figure.

randomized; every participant saw each of the 32 policies and 12 individuals only once during the experiment.

On the mystery choice trial, participants saw two colored boxes with question marks representing two unknown policy positions (Figure 3B). Underneath the two boxes were photos of Agents A and B and blue arrows pointing to their respective choices. We told participants that the mystery boxes contained Agent A's and Agent B's preferred political policy stance. Participants had to indicate which one of the two unknown policy positions they would rather choose (e.g., "Which would you choose? Remember, Jenny and Joyce know what's inside the boxes."). Thus, participants had to infer which one of the two mystery policy positions they would rather choose based on the choices of the two agents. The response table summarizing participants' and all agents' preferences during the block was still visible during the mystery choice trial. After the mystery choice trial, participants started another block with a new set of three agents and a new set of eight policy positions. All experimental materials can be found on the OSF repository, linked above.

We instituted timed delays because the task required reading comprehension of the policies and we wanted to ensure that participants were encoding the policies as well as each agent's preferences. Specifically, participants could only make a self-choice after a 2-s delay, a prediction for an agent after a 1-s delay, and a mystery choice after a 2-s delay. Additionally, feedback on predictions for other agents' preferences was displayed for 1 s.

Participants completed four blocks in total: two wherein Agent C agreed with the participant on 87.5% of trials (high C-agreement) and two wherein Agent C agreed with the participant only on 12.5% of the trials (low C-agreement). In each block, Agents A and B each agreed with the participant on half of the trials, and Agent C agreed with Agent B on 62.5% of the trials. Only Agents A and B were shown during the mystery choice trial.

After completing the four blocks (two high C-agreement blocks and two low C-agreement blocks), participants were probed for comments about the task and asked for their political affiliation (if any). They then completed four multiple-choice questions from the five-question American National Election Studies political engagement survey (Carpini & Keeter, 1993; "Which party holds the majority in the House" [Democratic/Republican/Not Sure], "Who decides if a law is constitutional?" [President/Congress/Supreme Court/Not Sure], "Which one of the parties is more conservative?" [Democratic/Republican/Not Sure], "What majority share is needed for the U.S. Congress to override a Presidential veto?"[One-half/Two-thirds/Three-fourths/Nine-tenths/Not Sure]). We used these questions to ensure that participants in our final pool were paying attention to the task and minimally engaged with political issues more generally.

Results

Across all four experiments, we used a logistic regression predicting whether participants chose Agent B's choice during the mystery choice trial as a function of Agent C's agreement (high: 87.5% vs. low: 12.5%) with the participant. Given the repeated-measures design of the experiment, we included random slopes by block for each participant to account for block order and subject effects. The model indicated a significant difference between the high and low C-agreement condition blocks predicting the probability of choosing Agent B's policy position on the mystery choice trial, $b = 0.348$, Wald's $z = 2.05$, 95% CI [0.015, 0.682], $p = .041$. As predicted, participants were more likely to choose Agent B's policy position on the mystery choice trial after a high C-agreement block ($M = 53.24\%$, $SD = 49.98$), which organized the participant, Agent B, and Agent C into a latent group, compared with a low C-agreement block ($M = 44.97\%$, $SD = 49.83$). Thus, people were more likely to align themselves with Agent B relative to Agent A when Agent C created a latent group including the participant and Agent B, despite the fact that both Agents A and B shared 50% of their preferences with the participant.

Discussion

We replicated the results of Gershman et al. (2017) in the more consequential domain of politics. Some of the political issues featured in Experiment 1 were, however, somewhat obscure, potentially rendering any given agent's agreement or disagreement with participants meaningless. Thus, a more stringent test of the effect of latent group structures would be one based on preferences that are more central to participants' political identities.

Experiment 2: Latent Group Structure Based on High-Consensus Political Issue Preferences Drives Choice

Here, we sought to replicate the findings of Experiment 1 with high-consensus political issue preferences (i.e., those issues with which most of the population agrees or disagrees). This approach makes it more likely that any one agent expresses stereotypically incoherent preferences (e.g., Agent B is both prochoice and for the death penalty), increasing the likelihood that participants would rely on similarity to self on specific issues rather than on latent group structures. As such, Experiment 2 is a more conservative replication of Experiment 1.

Method

Participants and exclusions. As with Experiment 1, we aimed for a total of 150 participants after exclusions; we recruited 168 participants on AMT. We excluded five people for not identifying as either male or female (1 person), for commenting that they did not find the task believable (1 person), or for failing all four questions of the political engagement survey (3 people). This left us with a sample size of 163 participants (84 female, $M_{age} = 35.77$ years, $SD = 11.66$). We also excluded 30 mystery choice trials in which the reaction time (RT) was longer than 20s, leaving us with 622 trials (310 high C-agreement, 312 low C-agreement).

We also screened participants to ensure they had not participated in Experiment 1. When AMT workers attempted to accept the task, we checked their worker IDs against the list of workers who completed Experiment 1. Workers found on the list were not allowed to complete Experiment 2, and workers who completed Experiment 2 were likewise logged for future prescreens. After data collection, we also checked the list of worker IDs across both experiments and found that there was no overlap between the two subject pools. This method of preventing AMT workers from participating in multiple, related studies was used across all our experiments here to ensure that the subject pools for each experiment did not contain the same workers.

Materials. We again used ISideWith.com to select the political issues, but this time we included 32 issues that that had accumulated at least 500,000 votes and had the *greatest* agreement/disagreement discrepancies as of March 2017. We also selected new images to represent the new issues (see OSF for complete materials).

Procedure. The entire experiment was identical to Experiment 1 except for the political issues on which participants' and agents' preferences were based.

Results

To model the probability of choosing Agent B's choice in the mystery trial as a function of latent group structure, we used a logistic regression including random slopes and intercepts to account for block order and subject effects, respectively. The model indicated a significant difference between the high and low C-agreement condition blocks predicting the probability of choosing Agent B's policy position on the mystery choice trial, $b = 0.482$, Wald's $z = 2.952$, 95% CI [0.162, 0.802], $p = .003$. Specifically, participants were more likely to choose Agent B's policy position on the mystery choice trial after a high C-agreement block ($M = 55.16\%$, $SD = 49.81$), which organized the participant, Agent B, and Agent C into a latent group, compared with a low C-agreement block ($M = 43.27\%$, $SD = 49.62$).

Discussion

Replicating Experiment 1, participants were more likely to choose Agent B's mystery trial choice when Agent C formed a latent group with the participant and Agent B, compared with when Agent C did not form such a group, even though both Agents A and B shared 50% of their preferences with the participant. Thus far, our data are more consistent with a latent group structure learning rather than a dyadic similarity account.

Experiment 3: Latent Group Structure Drives Choice and Trait Attribution

Here we sought to replicate Experiment 2 and examine whether inferred groups influence judgments beyond choice. Specifically, we asked participants to rate the agents in terms of likability, morality, and competence. If groups exert promiscuous influence, then the agent belonging to the participant's own inferred group should be rated higher along these positive dimensions compared with the agent belonging to a different group, but only when latent groups are formed. We chose these traits due to their centrality to social cognition and impression formation (Fiske, Cuddy, & Glick, 2007; Fiske, Cuddy, Glick, & Xu, 2002; Goodwin, Piazza, & Rozin, 2014). The preregistration of Experiment 3 is also available at OSF: https://osf.io/9s2hv/

Method

Participants and exclusions. We aimed for 250 participants after exclusions in order to have sufficient power to detect a small effect, so we recruited 293 participants via AMT. We excluded 10 participants for not identifying as male or female (1 person) or for failing all four questions of the political questionnaire (9 people). This left us with a sample size of 283 participants (135 female, $M_{age} = 34.57$ years, $SD = 10.37$). We also excluded 98 mystery choice trials (out of 1132 trials) in which participants took longer than 20 s to respond. This left us with 1,034 valid trials (512 high C-agreement, 522 low C-agreement).

Materials. We used the same stimuli as in Experiment 2.

Procedure. The entire experiment was identical to Experiment 2 except for the addition of trait judgments at the end of each block. Thus, within the main task, each block consisted of three phases: (a) eight regular trials during which participants expressed their own preferences and learned about others' political preferences, (b) a mystery choice trial, and (c) three trait judgments of each of the other individuals in the block.

After the mystery choice trial, we presented the picture of one of the agents and asked participants to judge how likable, moral, and competent that agent was (e.g., "How *likeable* is this person?") on a 9-point Likert scale (1 = *Not at All* to 9 = *Extremely*). Participants made these ratings sequentially and repeated this process until they had made three trait judgments for each of the three agents shown in the block. We randomized the order of the trait questions for each agent as well as the order of the agents. After completing the trait judgments, participants started another block with a new set of three agents and a new set of eight policy positions.

Results

Probability of choosing Agent B's choice in the mystery trial. To model the probability of choosing Agent B's choice in the mystery trial as a function of latent group structure, we used a logistic regression including random slopes and intercepts to account for block order and subject effects, respectively. Replicating Experiments 1 and 2, we found a significant difference between the high and low C-agreement blocks in the percentage of trials in which the participant chose Agent B's mystery stance, $b = 0.352$, Wald's $z = 2.78$, 95% CI [0.104, 0.600], $p = .005$. Participants were more likely to choose Agent B on the mystery choice trial after a high C-agreement block ($M = 52.54\%$, $SD = 49.98$), which organized the participant, Agent B, and Agent C into a latent group, compared with a low C-agreement block ($M = 43.87\%$, $SD = 49.67$).

Trait judgments. To determine whether there was a difference in trait judgments of Agents A and B as a function of latent group structure, we regressed the trait ratings against four categorical variables: the type of rating (moral, competent, or likable), the block condition (high or low C-agreement), the agent (Agent A or Agent B), and whether the agent was the chosen agent in the mystery choice trial (chosen or nonchosen). We also included random slopes and intercepts to account for block order and subject effects, respectively. We began with a saturated model including all possible main effects and interactions, then eliminated all higher order interactions that did not pass significance threshold (i.e., where $p > .05$) and then compared the saturated model with the simpler model. We did this iteratively until likelihood ratio tests comparing the models indicated we were generating a significantly poorer model by removing any additional fixed effects. Our final model included all four main effects, and one two-way interaction (see below). We calculated the degrees of freedom using Satterthwaite's approximation as implemented in the lmerTest package in R and conducted simple effects analyses on least squares means computed from the omnibus model to

maintain experiment-wide error (Kuznetsova, Brockhoff, & Christensen, 2017).

The model indicated significant main effects of being the chosen agent, $F(1, 5661.2) = 601.15$, $p < .001$ as well as the type of rating ($F(2, 5661.2) = 30.71$, $p < .001$; see the online supplementary materials for detailed results). These main effects were qualified by a predicted Block Condition × Agent interaction, $F(1, 2881.8) = 26.14$, $p < .001$ controlling for which agent was chosen. After high C-agreement blocks, participants rated Agent B as more moral, competent and likable ($M = 5.56$, $SE = 0.073$) than Agent A ($M = 5.36$, $SE = 0.073$), mean difference = 0.208, 95% CI = [0.122, 0.294], $t(5661.2) = 4.74$, $p < .001$. On the contrary, after low C-agreement blocks, participants rated Agent B as less moral, competent, and likable ($M = 5.54$, $SE = 0.073$) than Agent A ($M = 5.65$, $SE = 0.073$), mean difference = −0.107, 95% CI [−0.195, −0.024], $t(5661.2) = -2.50$, $p = .013$ (Figure 4).

Discussion

First, we replicated the choice results of Experiments 1 and 2. Second, we found that even when participants did not choose Agent B's choice on mystery choice trials following high C-agreement blocks they rated Agent B as being more moral, competent, and likable compared with Agent A. That is, above and beyond whether or not the agent being rated had been previously chosen in the mystery choice trial, Agent B was rated more highly compared with Agent A. We observed the opposite pattern in low C-agreement blocks. This suggests that our effects are not driven by a dissonance account in which choice behavior drove evaluations. We do, however, note that the differences in the ratings of the two agents are driven mostly from movement in the ratings of Agent A across conditions. That said, the absolute values are not readily interpretable on this rating scale; as such we focus on the relative differences in trait ratings across conditions.

Thus, participants' latent group structures generalize to inform their trait attributions. What remains unresolved is whether participants fail to learn or use latent group structures when they have explicit, contradictory category group membership information.

Experiment 4: Latent Group Structures Drive Choice Even in the Presence of Category Labels

Here we sought to replicate the effect of latent groups from Experiments 1–3 in the presence of explicit category labels (i.e., random assignment to the orange or purple team) that directly contradicted the structure of the latent groups. Specifically, we designed the experiment so that the latent group member (i.e., Agent B in the high C-agreement condition) was always on the opposite color team. This approach allowed us to test whether participants would cease to use latent structure information in the presence of overt category labels (i.e., whether the effect of Agent C's choices on preference for B over A disappeared).

Method

Participants and exclusions. We included a low/high consensus between-subjects factor in this design (see below) and aimed for a total of 150 participants per consensus condition after exclusions (i.e., 300 people), so we recruited 338 participants on AMT. We excluded seven participants for either not identifying as either male or female (1 participant) or for failing all four questions of the political engagement survey (6 participants). At the end of the task, eight participants reported not being able to distinguish between the colors used in the task (see the Procedure section), and we excluded them from all analyses as well. All participants correctly identified their team assignment at the end of the task. This left us with 323 participants (153 female, M_{age} = 35.69 years, $SD = 11.44$). We excluded 76 mystery choice trials

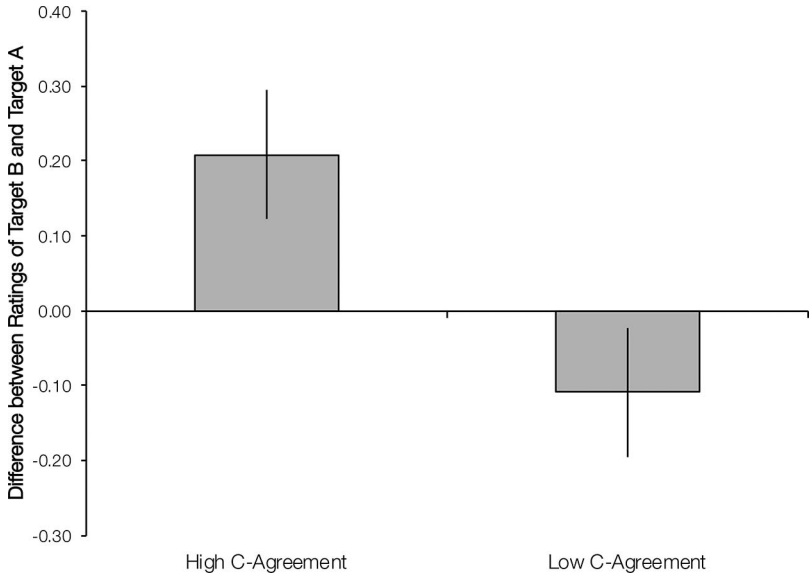

Figure 4. Difference in trait judgments for Agents B and A after high C-agreement blocks and low C-agreement blocks. Error bars represent 95% confidence interval.

(from a total of 1292) in which the RT was longer than 20 s, leaving us with 1,216 trials (587 high consensus: 293 high C-agreement, 294 low C-agreement; 629 low consensus: 313 high C-agreement, 316 low C-agreement).

Materials. Other than the use of orange- and purple-colored frames indicating agents' team membership, the materials were identical to Experiments 1 and 2.

Procedure. Participants were again recruited under the pretense of playing a game in which they would tell us about their political stances and learn about others' political issue preferences. They provided demographic information (age, ethnicity, and gender), after which we randomly assigned them to one of two teams (purple or orange). Following team assignment, we asked participants to correctly identify their team. If they answered incorrectly, participants saw the team assignment screen again and were asked to identify their team again. Participants repeated this process until they could correctly identify their team assignment.

Because the latent group effects were slightly weaker in Experiment 1 relative to Experiment 2, we also tested whether the effect of latent group preferences in the presence of category labels would be moderated by low-consensus versus high consensus political issues. We randomly assigned participants to see either the political issues from Experiment 1 (low consensus issues) or Experiment 2 (high consensus issues). The task was identical to Experiments 1 and 2, except during the mystery choice trial, we placed colored boxes (either orange or purple) around the agents' pictures and told participants that the boxes represented the team memberships of the agents. Agent B was always a member of the opposite team with respect to the participant, and Agent A was always on the same team as the participant. To decrease the multitude of colors on the mystery choice trials, we used gray mystery boxes and gray arrows (Experiments 1–3 used blue and green mystery boxes and light blue arrows).

After completing the general task, we asked participants to identify their team again and whether they had trouble distinguishing between colors ("Did you have any trouble distinguishing colors in this task?"). After this, participants entered comments (if any), identified their political party affiliation, and completed the same four questions from the political engagement survey included in Experiments 1–3.

Results

To model the probability of choosing Agent B's choice in the mystery trial as a function of latent group structure, we used a logistic regression in which we regressed the dummy variable of choosing Agent B against two categorical variables denoting political issue consensus condition (high or low) and C-agreement condition (high or low). As in the previous experiments, we also included random slopes and intercepts to account for block order and subject effects, respectively. The addition of an interaction term did not improve model fit; therefore, we did not include it in the final model.

The model indicated only a main effect of high versus low C-agreement blocks in the percentage of mystery choice trials in which the participant chose Agent B's political stance, $b = 0.486$, Wald's $z = 3.99$, 95% CI [0.247, 0.725], $p < .001$. Replicating Experiments 1–3, participants were more likely to choose Agent B's political stance after high C-agreement blocks ($M = 45.95\%$, $SE = 2.17$) compared with after low C-agreement blocks ($M = 34.35\%$, $SE = 2.07$; Figure 5). Note, however, that even in the high C-agreement condition, participants chose Agent B (the latent group, but different color team member) less than 50% of the time, on average.

Discussion

Participants were more likely to choose Agent B's choice when Agent C formed a latent group with the participant and Agent B compared with when Agent C formed no such group. More im-

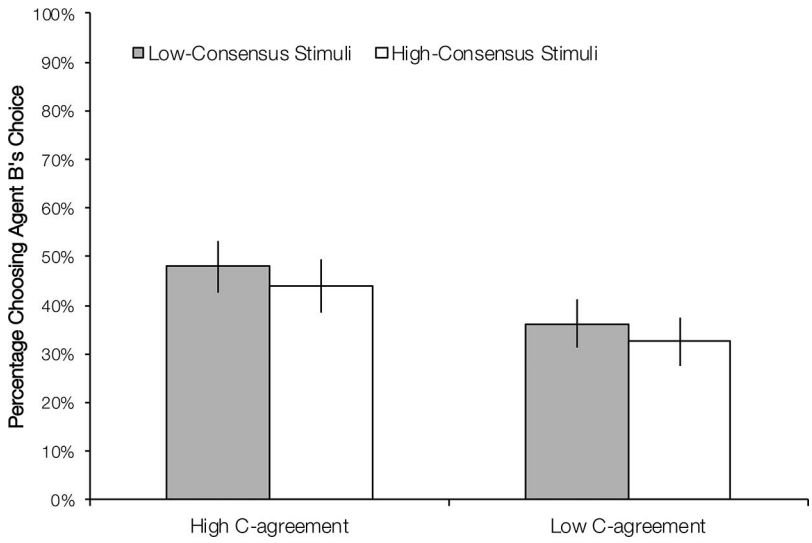

Figure 5. Mean percentage choosing Agent B on mystery choice trial for the high C-agreement and low C-agreement blocks across low consensus and high consensus stimuli. Error bars represent 95% confidence interval.

portantly, this effect persisted even though Agent B was always an outgroup member (i.e., on the other-color team).

General Discussion

Social categorization is a core social capacity that draws on many cognitive processes—matching to categories, mental state inferences, self-reference, and so forth (Cikara & Van Bavel, 2014). Social psychologists have argued that it is distinct from other forms of categorization in that we do not just sort people into categories (e.g., black/white), we sort them into ingroups and outgroups (i.e., mine/not mine), which are egocentrically defined. Nevertheless, we find that social categorization-driven effects adhere to the same principles as nonsocial category learning, particularly when people have to learn about group boundaries from other agents' behavior as opposed to labels. Of course, which category is salient—race, gender, profession—is highly context dependent (Turner et al., 1994). For example, assigning people to teams consisting of racial ingroup and outgroup members reduces race-based bias because the participants reorient themselves along these team dimensions rather than along racial dimensions (Kurzban, Tooby, & Cosmides, 2001). Given that social categorization is such a flexible and dynamic process, one open area of inquiry is how people accumulate group structure information from their immediate environments, especially in the absence of overt cues to peers' group membership. Latent structure learning is very well suited to grapple with the challenge of updating group boundaries efficiently.

In contrast to models where latent social groups are learned either solely via matching to group labels and stereotypes or dyadic similarity to the self, our results suggest that people take the relationships among agents into account when building representations of social groups. First we found that the degree to which individuals were willing to align with one of two agents was affected by the presence of a third agent, even in social domains with clear boundaries—political preference-based coalitions—that reliably drive discrimination and bias. If participants formed inferences about ingroup membership solely from dyadic similarity to the self or from perceived similarity to existing stereotypes, then forced-choice preference between any two agents should have remained unaffected by the presence of additional information about third agents. Furthermore, we found that people made use of the latent structures even when there were clear cues to alternate group boundaries (though they did not override the explicit team labels). Finally, we demonstrated that the latent social groups inferred in these tasks generalized to other judgments—specifically, modifying unrelated judgments of agents' competence, morality, and likability.

Of course, we are not the first to explore how the relationship between a participant and an agent may affect the perception of a third agent: see, for example, Heider's (1958) balance theory. Our model, however, may be the formal account by which balance theory operates. Moreover, applying a structure learning model, we can make specific predictions about the probability that a participant will like Agent B after X number of observations of Agent B's and Agent C's preferences.

Using the model, we can also predict a disidentification-driven bias in social evaluation and choice. Specifically, the model predicts a preference *against* Agent B after low C-agreement blocks and a preference *for* Agent B after high C-agreement blocks. Indeed, across all of our experiments, we find that participants are more likely to choose Agent A after low C-agreement blocks than they are to choose B after high C-agreement blocks. There is not explicit disidentification built into the model, except in the sense that decreasing the probability of one cluster necessarily raises the probability of other clusters (though this may not arise under different observed choice distributions). Nevertheless, the model predicts this disidentification pattern under these conditions—yet another phenomenon that this approach might help to explain.

Limitations and Future Directions

Although explicit grouping cues were pitted directly against latent groups in our experiment, we believe that these cues work together in the real world. Explicit groups may be impoverished or noisy indicators of preferences and values, and therefore latent groups are useful means by which richer and more accurate group structures can be inferred. This kind of synergy could be modeled by treating explicit groups as priors over latent groups, such that latent groups tend to resemble explicit groups, but can represent more complex structures if the observed data provide sufficient evidence.

Another limitation of our experiments is that the source of "promiscuous generalization" from preferences to traits is unclear. Why should people show this generalization, and what are its boundary conditions? One possibility is that latent groups are much more general than preferences, with a variety of other personality dimensions encompassed by the group structure. Experiments testing a broader array of such dimensions will be necessary to address this issue.

A third limitation is that the current work stops short of articulating the psychological processes that drive this clustering. Perhaps high similarity with Agent C increases the likelihood that participants adopt Agent C's "perspective" which then drives preferences for B over A (or in the low-C agreement condition, negative evaluations of C may taint B by virtue of their association, thereby driving preferences for A over B). Future work will explore whether this social clustering is driven by mentalizing, evaluation by association, or other candidate processes.

In conclusion, our results suggest that latent groups may be much more powerful than our previous work (Gershman et al., 2017) suggested: they extend beyond observed preferences, and they apply to socially consequential preferences even in the face of countervailing explicit labels. Thus, it is critical to understand the principles determining how such groups are inferred, their flexibility, and their generality.

Context

Social categorization is a core social capacity that draws on many cognitive processes. We introduce a novel mechanism which we believe likely operates in tandem with other social cues to guide social categorization and may provide a formal account of major theories of social relations, including Heider's

balance theory and disidentification. It is our hope that this approach will open new and exciting experimental avenues: for example, manipulating the amount of data and the reliability of different grouping cues which should lead to systematic differences in the degree to which explicit or latent groups dominate choice behavior.

References

Allport, F. H. (1954). The structuring of events: Outline of a general theory with applications to psychology. *Psychological Review, 61,* 281–303. http://dx.doi.org/10.1037/h0062678

Austerweil, J. L., Gershman, S. J., Tenenbaum, J. B., & Griffiths, T. L. (2015). Structure and flexibility in Bayesian models of cognition. In J. R. Busemeyer, J. T. Townsend, Z. Wang, & A. Eidels (Eds.), *Oxford handbook of computational and mathematical psychology* (pp. 187–208). New York, NY: Oxford University Press.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues, 55,* 429–444. http://dx.doi.org/10.1111/0022-4537.00126

Brewer, M. B. (2008). Depersonalized trust and ingroup cooperation. In J. Krueger (Ed.), *Rationality and social responsibility* (pp. 215–232). New York, NY: Psychology Press.

Byrne, D., & Nelson, D. (1965). Attraction as a linear function of proportion of positive reinforcements. *Journal of Personality and Social Psychology, 1,* 659–663. http://dx.doi.org/10.1037/h0022073

Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science, 3,* 14–25. http://dx.doi.org/10.1002/bs.3830030103

Carpini, M. X. D., & Keeter, S. (1993). Measuring political knowledge: Putting first things first. *American Journal of Political Science, 37,* 1179–1206. http://dx.doi.org/10.2307/2111549

Chang, L. W., Krosch, A. R., & Cikara, M. (2016). Effects of intergroup threat on mind, brain, and behavior. *Current Opinion in Psychology, 11,* 69–73. http://dx.doi.org/10.1016/j.copsyc.2016.06.004

Cikara, M., & Van Bavel, J. J. (2014). The neuroscience of intergroup relations: An integrative review. *Perspectives on Psychological Science, 9,* 245–274. http://dx.doi.org/10.1177/1745691614527464

Cikara, M., Van Bavel, J. J., Ingbretsen, Z. A., & Lau, T. (2017). Decoding "us" and "them": Neural representations of generalized group concepts. *Journal of Experimental Psychology: General, 146,* 621–631. http://dx.doi.org/10.1037/xge0000287

Elsbach, K. D., & Bhattacharya, C. B. (2001). Defining who you are by what you're not: Organizational disidentification and the National Rifle Association. *Organization Science, 12,* 393–413. http://dx.doi.org/10.1287/orsc.12.4.393.10638

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

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82,* 878–902. http://dx.doi.org/10.1037/0022-3514.82.6.878

Fiske, S. T., & Ruscher, J. B. (1993). Negative interdependence and prejudice: Whence the affect. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping: Interactive processes in group perception* (pp. 239–268). San Diego, CA: Academic Press.

Gelman, S. A., Collman, P., & Maccoby, E. E. (1986). Inferring properties from categories versus inferring categories from properties: The case of gender. *Child Development, 57,* 396–404. http://dx.doi.org/10.2307/1130595

Gershman, S. J., Pouncy, H. T., & Gweon, H. (2017). Learning the structure of social influence. *Cognitive Science, 41,* 545–575. http://dx.doi.org/10.1111/cogs.12480

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology, 106,* 148–168. http://dx.doi.org/10.1037/a0034726

Heider, F. (1958). *The psychology of interpersonal relations*. New York, NY: Wiley. http://dx.doi.org/10.1037/10628-000

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology, 53,* 575–604. http://dx.doi.org/10.1146/annurev.psych.53.100901.135109

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology a social identity perspective on polarization. *Public Opinion Quarterly, 76,* 405–431. http://dx.doi.org/10.1093/poq/nfs038

Iyengar, S., & Westwood, S. J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science, 59,* 690–707. http://dx.doi.org/10.1111/ajps.12152

Kinzler, K. D., Shutts, K., Dejesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition, 27,* 623–634. http://dx.doi.org/10.1521/soco.2009.27.4.623

Kubota, J. T., Banaji, M. R., & Phelps, E. A. (2012). The neuroscience of race. *Nature Neuroscience, 15,* 940–948. http://dx.doi.org/10.1038/nn.3136

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 98,* 15387–15392. http://dx.doi.org/10.1073/pnas.251541498

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82,* 1–26.

Lelkes, Y., & Sniderman, P. M. (2016). The ideological asymmetry of the American party system. *British Journal of Political Science, 46,* 825–844. http://dx.doi.org/10.1017/S0007123414000404

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47,* 1122–1135. http://dx.doi.org/10.3758/s13428-014-0532-5

Motyl, M., Iyer, R., Oishi, S., Trawalter, S., & Nosek, B. A. (2014). How ideological migration geographically segregates groups. *Journal of Experimental Social Psychology, 51,* 1–14. http://dx.doi.org/10.1016/j.jesp.2013.10.010

Patton, C. (1995). Refiguring social space. In S. Seidman & L. Nicholson (Eds.), *Social postmodernism: Beyond identity politics* (pp. 216–249). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511520792.010

Rhodes, M., & Chalik, L. (2013). Social categories as markers of intrinsic interpersonal obligations. *Psychological Science, 24,* 999–1006. http://dx.doi.org/10.1177/0956797612466267

Rhodes, M., Leslie, S. J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences of the United States of America, 109,* 13526–13531. http://dx.doi.org/10.1073/pnas.1208951109

Rokeach, M., Smith, P. W., & Evans, R. I. (1960). Two kinds of prejudice or one. In M. Rokeach (Ed.), *The open and closed mind* (pp. 132–168). New York, NY: Basic Books.

Sherif, M. (1966). *In common predicament: Social psychology of intergroup conflict and cooperation*. Boston, MA: Houghton Mifflin.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1,* 149–178.

Taylor, M. G., & Gelman, S. A. (1993). Children's gender-and age-based categorization in similarity and induction tasks. *Social Development, 2,* 104–121. http://dx.doi.org/10.1111/j.1467-9507.1993.tb00006.x

Terry, D. J., & Hogg, M. A. (1996). Group norms and the attitude-behavior relationship: A role for group identification. *Personality and Social Psychology Bulletin, 22,* 776–793. http://dx.doi.org/10.1177/0146167296228002

Turner, J. C., Oakes, P. J., Haslam, S. A., & McGarty, C. (1994). Self and collective: Cognition and social context. *Personality and Social Psychology Bulletin, 20,* 454–463.

Zhong, C. B., Galinsky, A. D., & Unzueta, M. M. (2008). Negational racial identity and presidential voting preferences. *Journal of Experimental Social Psychology, 44,* 1563–1566. http://dx.doi.org/10.1016/j.jesp.2008.08.001

Zhong, C. B., Phillips, K. W., Leonardelli, G. J., & Galinsky, A. D. (2008). Negational categorization and intergroup behavior. *Personality and Social Psychology Bulletin, 34,* 793–806. http://dx.doi.org/10.1177/0146167208315457

Received August 15, 2017
Revision received May 9, 2018
Accepted May 10, 2018 ■