

## **Transmission of societal stereotypes to individual preferences through social reinforcement learning**

B. S. Stillerman<sup>1</sup>, B. R. Lindström<sup>2</sup>, D. T. Schultner<sup>3</sup>, L. M. Hackel<sup>4</sup>, D. R. Hagen<sup>3</sup>, N. B. Jostmann<sup>3</sup>,  
& D. M. Amodio<sup>\*1, 3</sup>

<sup>1</sup>New York University

<sup>2</sup> VU University Amsterdam

<sup>3</sup>University of Amsterdam

<sup>4</sup>University of Southern California

Corresponding author:

David M. Amodio  
Department of Psychology  
University of Amsterdam  
Nieuwe Achtergracht 129, REC G  
1001 NK Amsterdam, NL  
david.amodio@gmail.com

Keywords: Prejudice, stereotypes, learning, social, reinforcement

**Abstract**

How are societal stereotypes transmitted to individual-level group preferences? We propose that exposure to a stereotype, regardless of whether one agrees with it, can shape how one experiences and learns from interactions with members of the stereotyped group, thereby inducing individual-level group preferences. In a series of experiments, participants interacted with players from two groups, described with either positive or negative stereotypes, in a reinforcement learning task presented as a money sharing game. Although players' sharing behaviors were equated between groups, participants formed more positive reward associations with players from positively-stereotyped than negatively-stereotyped groups. Computational modeling revealed that this preference was due to a stereotype effect on both priors regarding group members' behavior and learning rates from player feedback. We then show that these stereotype-induced preferences, once formed, spread unwittingly to others who observe these interactions. Finally, we show that this pattern of biased learning and propagation occurs even when the stereotypic message comes from an opposed politician, suggesting that this transmission—from societal stereotypes to personal preferences—can occur regardless of one's intentions. These findings offer a mechanistic account of how exposure to social stereotypes can transform into personal group-based preferences and spread to others.

When politicians refer to members of an ethnic group as “criminals and drug dealers,” as Donald Trump famously did during his 2016 presidential campaign announcement, people may dismiss the epithets as mere rhetoric.<sup>1</sup> Yet such messages, conveyed explicitly, are nevertheless encoded in the recipient’s memory, from where they might shape future experiences with group members. We asked whether exposure to societal stereotypes can influence the way people subsequently learn from members of the targeted group in direct interactions, such that it biases how they experience and learn from group members’ behavior to create individual-level prejudices, which may then spread to others.

Stereotypes are societally-held beliefs about a group that generalize to its members (Hilton & Von Hippel, 1996; Fiske, 1998). By providing expectancies of group members’ behaviors (Darley & Gross, 1983), stereotypes shape how we perceive, interpret, and judge a person’s actions (Devine, 1989; Kawakami et al., 2017; Kunda & Spencer, 2003). Much prior work shows that societal stereotypes are distinguishable from individual-level preferences; for example, most people have knowledge of societal stereotypes, yet many people consciously reject them (Devine & Elliot, 1995) and act to overcome them through direct learning about an individual group member (Fiske & Neuberg, 1990) or by exerting cognitive control (Devine, 1989).

Here, we considered an unexplored possibility: if stereotypes shape our expectancies in direct interactions, could they influence how we experience and thus learn about a group members? If so, this process would undermine the debiasing effects of individuated learning (because perceptions of individuating information would be biased) and control (because the presence of bias would be undetected). As a result, a perceiver may form a stereotype-consistent preference based on their biased perception of a group members’ behavior.

How can stereotype knowledge affect the way we learn about group members in direct social interactions? In direct interactions, a perceiver learns about a group member through an exchange of actions and feedback—a process characterized by instrumental learning (i.e., reward reinforcement; Hackel, Kogon, Amodio, & Wood, 2022; Kurdi, Gershman, & Banaji, 2019, Lindström, Selbing, Molapour, & Olsson, 2014). Unlike stereotype knowledge, represented in terms of concepts and expressed in judgments, instrumentally-learned preferences are formed incrementally through repeated interaction and feedback, encoded in terms of reward value (Daw, O’Doherty, Dayan, Seymour, & Dolan, 2006), and expressed in choice behaviors that reflect an individual’s preferences (Delgado, Frank, & Phelps, 2005; Shohamy & Wagner, 2008). Importantly, instrumental learning may be shaped by priors, which affect one’s expectations about feedback (Delgado, Frank, & Phelps, 2005; Doll, Jacobs, Sanfey, & Frank, 2009). If stereotypes function similarly to priors, then exposure to stereotype

---

<sup>1</sup> Donald Trump, in his speech launching of his 2015 presidential campaign, famously referred to Mexican immigrants to the US as criminals, drug dealers, and rapists.

messages may also bias how a reward association is updated in response to a group member's feedback, thereby producing a personal (i.e., individual-level) group-based preference.

Based on this analysis, we hypothesized that stereotype messages can induce personal (individual-level) group-based preferences through two concerted processes: First, exposure to a positive or negative stereotype sets initial expectations (i.e., *priors*) for a group member's behavior which, during direct interactions; second, stereotypes influence the degree to which reward representations of a group member are updated in response to feedback across repeated interactions (i.e., the *learning rate*), such that updating occurs differently for members of positively and negatively stereotyped groups. Whereas prior research has examined the role of biased sampling in the formation of stereotypes and evaluations (Allidina & Cunningham, 2021; Bai, Fiske, & Griffiths, 2022; Fazio, Eiser, & Shook, 2004), the current hypothesis, by contrast, asks how existing stereotype knowledge can influence the encoding of a group member's feedback, leading to the biased group preferences.

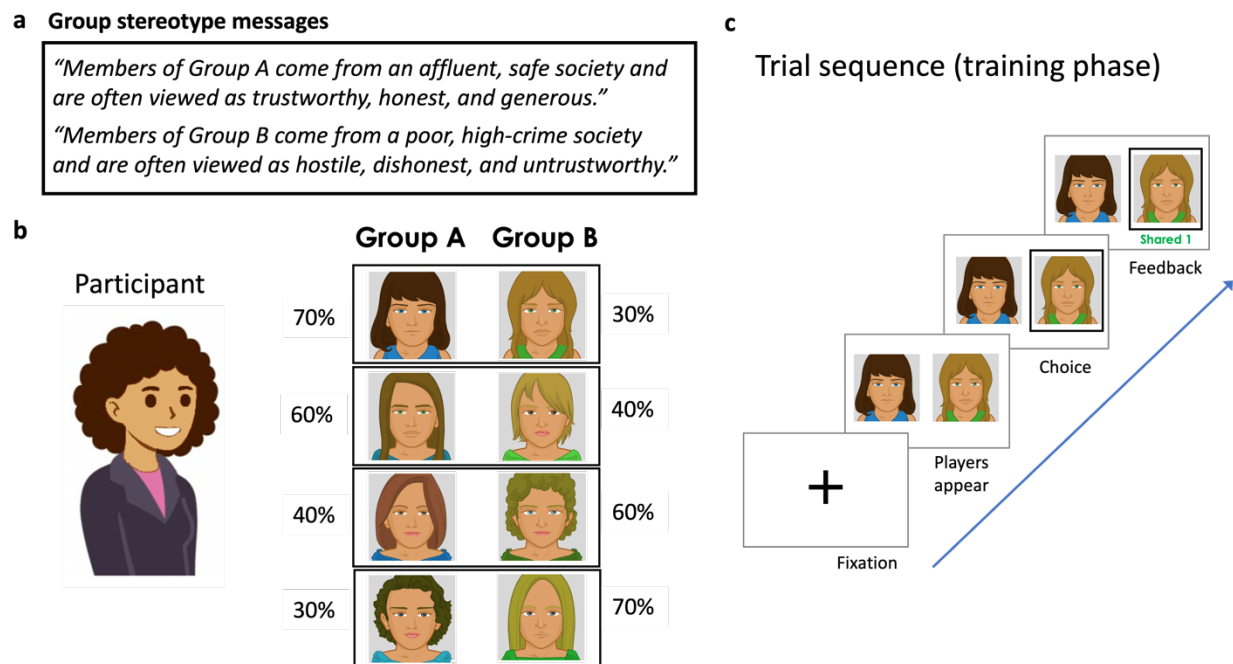
We tested this *biased learning* hypothesis across six experiments, in which we predicted that stereotype descriptions of groups would bias participants' instrumental learning during direct interactions with group members, even when participants reject the source of the stereotype. We examined this effect in participants' behaviors and tested our hypothesis using computational modeling, and then further examined how such biases, once acquired and expressed, may spread to others who observe these direct interactions.

In Experiments 1-3, participants interacted with people from two different social groups in an online game. These two groups were labeled as Group A and B (counterbalanced) in the task, ostensibly to maintain their anonymity, but described using positive or negative societal stereotypes of White and Black Americans, respectively (Devine & Elliot, 1995): Group A was characterized as coming from a relatively wealthy, safe, and highly educated community, whereas Group B's community was characterized as relatively poor and uneducated and with a high crime rate (see Supplementary Information [SI]). This approach allowed us to isolate effects of stereotypes on learning while controlling for participants' existing group knowledge. Despite these group descriptions, participants were told that individual group members varied in their tendency to share money and therefore, given participants' explicit goal to earn money, they should attend to the sharing rate of each player. Participants then completed a money sharing task with members of both groups, receiving cash payouts for their winnings.

The sharing game was adapted from a widely-used probabilistic reward reinforcement learning task (Frank, Seeberger, & O'Reilly, 2004). In this version, participants interacted with four players from each group. Within each group, each player shared money at a different fixed rate (70%, 60%, 40%, or 30%), but average sharing rates were equated between groups (Figure 1).

Participants first completed a training phase, in which they could learn from feedback on each trial. This was followed by a test phase, in which they made choices, based on what they learned, but without feedback.

On each round of training (160 trials), participants were presented with a preset pair of players—one from each group, with fixed complementary sharing rates (e.g., Players A and B)—and chose, via button press, with whom to interact (Figure 1). Reward feedback, displayed immediately beneath the image of the chosen player, indicated whether the chosen player shared (+1 or 0 points). Participants knew that only one player would share on each round.



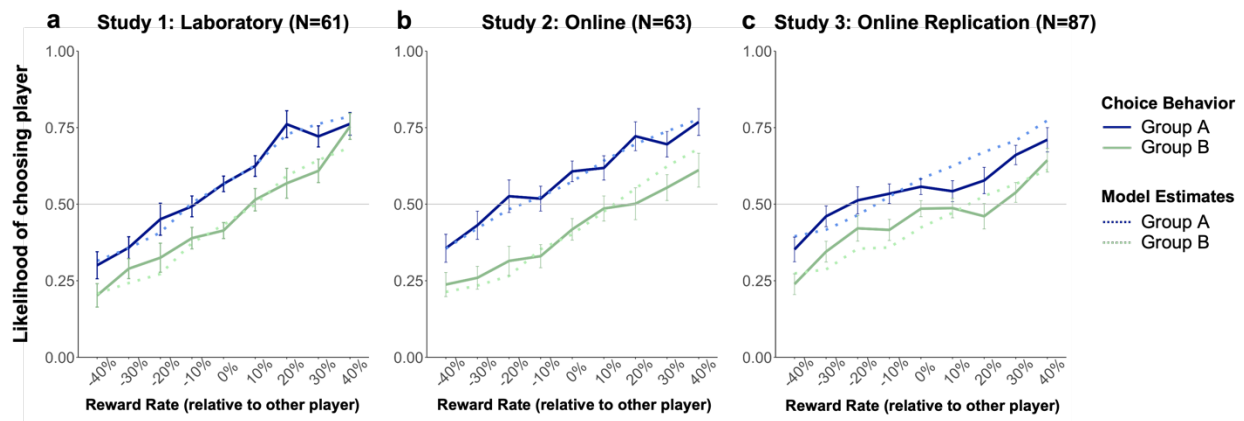
**Figure 1.** Schematic of the learning task training phase. **a**, Participants were exposed to group stereotype messages and then **b**, interacted with members of two groups, one described according to positive stereotypes, the other described according to negative stereotypes. Group labels, member features (e.g., hair, shirt color), and gender were counterbalanced across participants. **c**, In the training phase, participants chose between players (group members) with reciprocal reward rate (e.g., 70% and 30%) and received reward feedback, as shown in this sample trial. In the test phase (not shown), participants chose between all possible intergroup pairs of players (e.g., 70% and 70%) and received no feedback.

Following the training phase, participants completed the test phase (96 trials), in which they viewed and selected between all possible pairs of Group A and B members. This allowed us to assess participants' choices between novel pairs of players who shared at identical rates during training, in addition to every other pairing of sharing rate. Hence, the test phase provided a fine-grained behavioral assessment of learned reward associations with each member of the two groups (Frank et al., 2004). Although feedback was not provided, to prevent further

learning, participants were told they would receive payout for their choices following task completion.

## Results

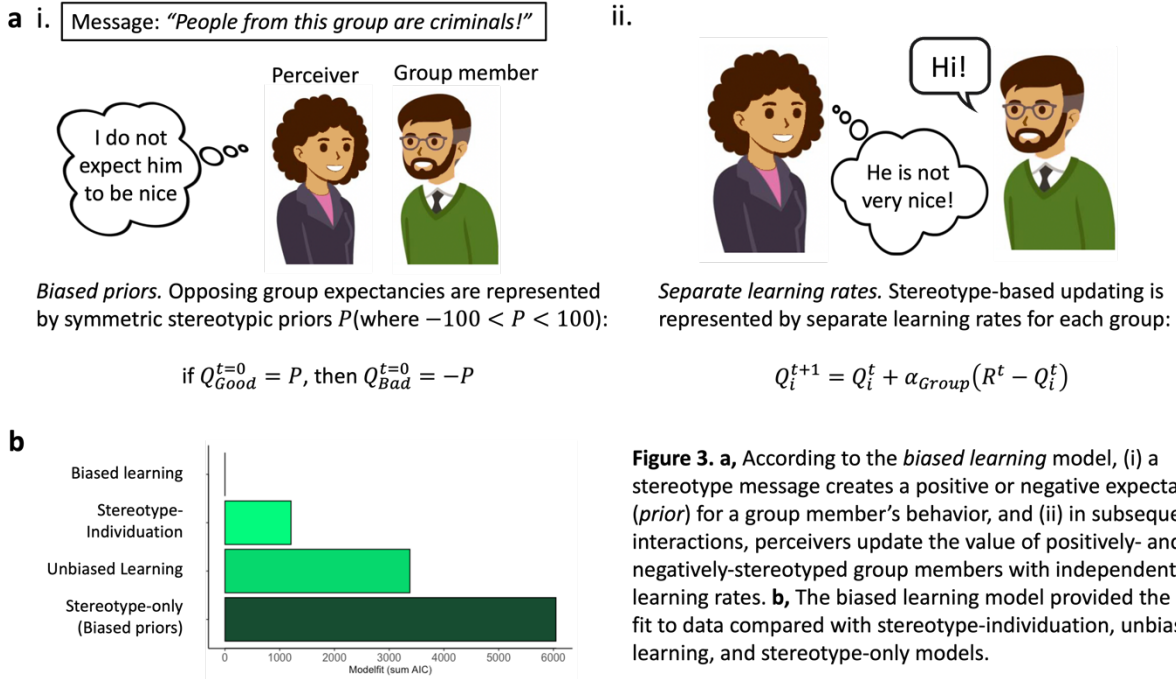
In Study 1 ( $N = 61$  laboratory participants; full data and code for all studies will be made available prior to publication), we tested whether stereotypic group descriptions influenced participants' choices of individual players, despite equivalent sharing rates between groups—the hallmark of group-based prejudice. Analysis of test phase behavior showed that while participants learned the general pattern of rewards, choosing players with higher sharing rates on average ( $B = 2.68$ ,  $SE = 0.19$ , Wald  $z = 14.43$ ,  $p < .001$ ; all tests are two-tailed), their choices were also significantly biased by players' group membership ( $B = 0.52$ ,  $SE = 0.06$ , Wald  $z = 9.33$ ,  $p < .001$ ; Figure 2a). That is, direct learning failed to overcome the influence of stereotypes. This effect of group membership emerged despite participants' extensive direct experience with players' actual sharing rates, which were equated between groups, and their monetary incentive to choose accurately (results were similar during the training phase; see SI).



**Figure 2.** Choice behavior for the test phase of Studies 1–3 (Panels a–c, respectively). Participants' choices (solid lines) demonstrated both successful learning of reward contingencies and a group bias. Reward rate, displayed on the x axis, represents the actual reward rate of a given player minus the actual reward rate of the alternative player in a trial. Error bars indicate standard error. Dotted lines show estimates from the biased learning model, which combined group-based priors and separate learning rates (see text for details).

While these regression analyses show that choice preferences were guided by the group stereotype as well as actual reward feedback, they do not address our specific hypothesis that this pattern involves an effect of stereotype knowledge on learning. To test this hypothesis, we fit behavior to a computational model specifying this process, adapted from Doll et al. (2009). We conceptualized stereotype effects on group expectancy as separate *priors* for positively and negatively stereotyped groups, which set participants' initial choice tendencies. Stereotype

effects on the updating of reward associations were represented by separate *learning rates* for positively- and negatively-stereotyped groups. Thus, according to this *biased learning model* (Figure 3), the behavioral effects of stereotypes on instrumental learning reflect a combination of divergent group priors and separate group learning rates.



We compared the biased learning model with alternatives representing other models of stereotyping and impression formation: (a) the classic stereotyping effect, whereby stereotypes affect judgment with no effect on learning (biased priors without learning), (b) the stereotype-individuation model (e.g., Fiske & Neuberg, 1990; Rothbart, 1981), in which new learning replaces the effect of stereotypes (biased priors and a single, unbiased learning rate), and (c) individuated learning with no effect of stereotypes (a single learning rate and no priors), in addition to other plausible reinforcement learning and Bayesian accounts (see SI for model specifications and results). Model comparisons indicated that the biased learning model, which included stereotypic priors and separate learning rates for each group, best explained behavior (see model fits in Figure 1a and SI).

This effect was replicated in two online experiments (Study 2:  $N = 62$ ; Study 3:  $N = 87$ ): In both, stereotypic group descriptions again significantly influenced participants' choice preferences in the training and test phases (Study 2 test phase:  $B = 0.79$ ,  $SE = 0.06$ , Wald  $z = 13.86$ ,  $p < .001$ ; Study 3 test phase:  $B = 0.48$ ,  $SE = 0.05$ , Wald  $z = 9.58$ ,  $p < .001$ ; see SI), despite equivalent average sharing rates between groups and participants' explicit goal to learn about individual players to maximize earnings. Study 3 also assessed participants' subjective estimates of player

sharing rates and compared them with behavioral choice preferences. Although correlated with choice behavior, these subjective reports could not fully account for the group bias in choice preferences; by contrast, subjective reward estimates predicted participants' deliberative responses to players in a trust game, suggesting a distinction between stereotype effects on behavior and subjective report (see SI).

Finally, replicating Study 1, computational modeling of Study 2 and 3 data showed that stereotypes affected both group-based priors and group learning rates (see SI). To examine the unique effects of stereotype-biased priors and learning rates on participants' preferences, we submitted estimates of priors and group-specific learning rates, derived from the biased-learning model, to a regression predicting group-based choice behaviors. This analysis combined Studies 1-3 data. Results indicated that the group bias in preferences reflected both stereotype-based priors and insufficient updating for the negatively-stereotyped group; that is, initial expectancies for the negatively-stereotyped group were lower, relative to the positive group, and were not sufficiently updated in response to group members' actual reward feedback (see SI). This pattern directly demonstrates an effect of stereotypes on learning, beyond their previously-shown effect on expectancy.

Having observed the transmission from societal stereotypes to individual group preferences in Studies 1-3, we next considered a secondary form of transmission, whereby stereotype-based preferences spread to people who merely observe interactions between a stereotype-exposed actor and group member. In Study 4 ( $N = 124$ , preregistered: [https://aspredicted.org/STK\\_EXP](https://aspredicted.org/STK_EXP)), participants played the money sharing game used in Studies 1–3. However, instead of learning directly from group members in a training phase, participants observed the training-phase choices and feedback of a prior participant (demonstrator) across 160 trials. Observers were told they should observe and learn each player's sharing rate to improve their own chances of winning money in a subsequent test phase with the same players. Observers were told only that players came from two different groups; they were not exposed to the stereotype descriptions provided to demonstrators. Each Study 4 participant (observer) viewed the learning phase interactions of a participant from Study 2 (in a yoked design), in which a demonstrator made choices and received feedback from players. Participants then made their own choices in a test phase (identical to the test phase in Studies 1-3). Following the task, participants reported estimated reward rates for each player, as in Study 3.

Did the mere observation of demonstrators' behavior and feedback induce a group preference in observers? It did: observers exhibited a significant group bias in their own test phase choices, despite having no direct experience with the stereotype ( $B = 0.32$ ,  $SE = 0.04$ , Wald  $z = 8.03$ ,  $p < .001$ ; Figure 4), and the magnitude of their group bias correlated with the degree of bias exhibited in the demonstrator's own test phase choices ( $B = 0.28$ ,  $SE = 0.09$ , Wald  $z = 3.21$ ,  $p =$

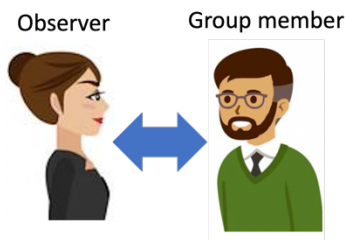


.001). This group bias remained after adjusting for Study 4 participants' subjective estimates of player reward rates ( $B = 0.26$ ,  $SE = 0.04$ , Wald  $z = 6.52$ ,  $p < .001$ ). These findings suggest a novel form of bias propagation, such that explicit stereotypes affect not only direct learners in social interactions, but also observers of these interactions who had not been exposed to the stereotype.

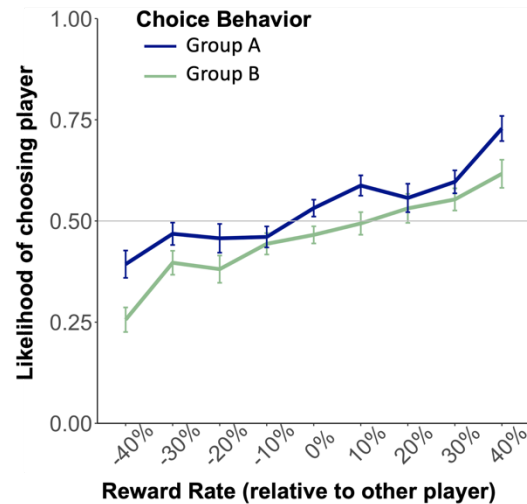
**a Training Phase:** Observer views yoked demonstrator interactions with group members



**b Test Phase:** Observer interacts directly with group members



**c Study 4: Observer choice behavior**



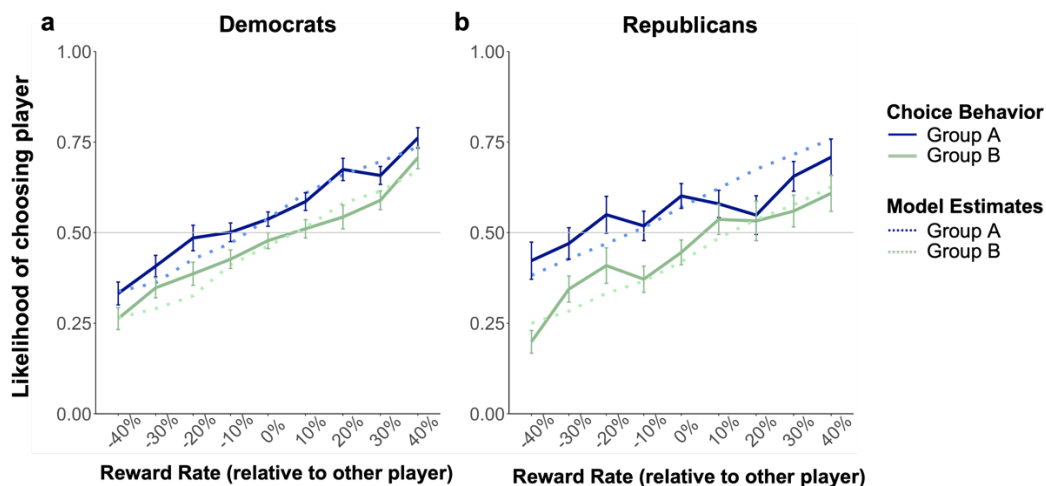
**Figure 4.** Observational learning task and results from Study 4: (a) Participants (observers) view the training phase interactions between group members and prior participants who had knowledge of stereotypes, (b) Observers then completed test phase choices with group members, (c) Results show that observers exhibited the past participants' group bias in their learning of reward rates. Reward rate, displayed on the x axis, represents the actual reward rate of a given player minus the actual reward rate of the alternative player in a trial. Error bars indicate standard error.

Finally, having found that social stereotypes can be internalized in one's own choice preferences through instrumental learning and propagated to others through observation, we returned to the question we began with: can a politician's stereotypical rhetoric about a social group become internalized, even among people who object to the politician and oppose their views? To test this question, we adapted the design of Study 1 to present the negative group description as a quote from U.S. President Donald Trump. As in Studies 1–3, American participants ( $N = 200$ , preregistration: [https://aspredicted.org/YJS\\_TFP](https://aspredicted.org/YJS_TFP)) learned they would complete a money sharing task with players who came from two different communities. However, in this study, they learned that President Trump described one group as "coming from a society that is economically poor, with a high rate of unemployment and serious crimes

such as robbery, assault, and murder” and as “hostile, untrustworthy, and dishonest.”

Participants read that President Trump had not commented on the second group, from which the other players came, such that it constituted a control group. Participants then completed training and test phases of the social reward learning task, as in Study 1 and, upon completion, reported their political party identification (Democrat,  $N = 162$  or Republican,  $N = 38$ ) and their agreement with Donald Trump’s positions (1= strongly disagree, 5=strongly agree). As expected, Democrats indicated less agreement with Trump ( $M = 1.98$ ,  $SD = 1.07$ ) than did Republicans ( $M = 3.33$ ,  $SD = 0.88$ ), Welch’s  $t(135.68) = 9.33$ ,  $p < .001$ , 95% CI = [1.07, 1.64].

Did the stereotypical group description attributed to Donald Trump influence participants’ choice preferences? Consistent with Studies 1-3, participants were more likely to avoid members of the group disparaged by Trump, despite the equivalent average sharing rates between groups ( $B = 0.40$ ,  $SE = 0.03$ , Wald  $z = 12.78$   $p < .001$ ). Although our sample included fewer self-identified Republicans than Democrats, our primary interest was in the behavior of Democrats: would they form group-based choice preferences despite their opposition to the messenger? Indeed, this effect was significant among both Democrats ( $N = 138$ ;  $B = 0.31$ ,  $SE = 0.04$  Wald  $z = 8.14$ ,  $p < .001$ ) and Republicans ( $N = 62$ ;  $B = 0.62$ ,  $SE = 0.06$ , Wald  $z = 10.83$ ,  $p < .001$ ), although an interaction indicated a relatively larger effect among Republicans ( $B = 0.32$ ,  $SE = 0.07$ , Wald  $z = 4.58$ ,  $p < .001$ ).



**Figure 5.** Group bias in choice preferences in Study 5 for (a) Democrat and (b) Republican participants following exposure to a Republican leader’s negative description of Group B. Participants of both political parties exhibited group bias, in addition to learning reward rates. Reward rate, displayed on the x axis, represents the actual reward rate of a given player minus the actual reward rate of the alternative player in a trial. Error bars indicate standard error. Dotted lines show predictions from the biased learning model, which combined biased group-based priors and learning rates (see text for details).

Computational modeling again indicated best fit to the biased learning model, reflecting different priors and independent learning rates for each group—a pattern also seen in separate tests for Democrats and Republicans (see SI). Furthermore, behavioral choice preferences were predicted by unique effects of priors as well as learning rates, replicated Studies 1-3.

In a final study, we tested whether the group bias, conveyed by a political leader and internalized by direct learners regardless of their political views, could spread to others through mere observation of a direct learner's behavior, as in Study 4. Participants in Study 6 observed prior training-phase choices of yoked Study 5 participants (demonstrators), without exposure to the original group descriptions or knowing the demonstrator's political identification. Again, as in Study 4, there was evidence of bias propagation: observers exhibited a significant group bias in their own test-phase behavior, preferring to choose players from the group described neutrally to Study 5 participants over the group disparaged by President Trump,  $B = 0.25$ ,  $SE = 0.03$ , Wald  $z = 7.98$ ,  $p < .001$ , in addition to the effect of players' actual sharing rates,  $B = 1.82$ ,  $SE = 0.10$ , Wald  $z = 17.43$ ,  $p < .001$ . This effect was moderated by demonstrators' political views,  $B = 0.15$ ,  $SE = 0.07$ , Wald  $z = 2.20$ ,  $p = .028$ : as expected, given the stronger direct learning effect among Republicans, the degree of group bias conveyed to observers by Republican demonstrators was larger ( $B = 0.35$ ,  $SE = 0.05$ , Wald  $z = 6.36$ ,  $p < .001$ ) than that of Democrat demonstrators ( $B = 0.21$ ,  $SE = 0.04$ , Wald  $z = 5.28$ ,  $p < .001$ ). Importantly, these results demonstrate that group preferences were conveyed by the behavior of both Democrat and Republican demonstrators, despite its Republican source.

## Discussion

We asked whether exposure to societal stereotypes can induce personal group-based preferences by shaping the way one learns about group members in direct interactions. Across four studies, we found that positive and negative group descriptions regarding stereotypes and social inequalities, conveyed explicitly, biased the process of reward reinforcement learning in direct interactions with group members. Computational modeling indicated this process involved the interplay of two processes: stereotypes set initial expectancies for each group and then biased updating of reward values of individual group members in direct interactions. These results showed that, in this context of interactive learning, participants did not merely apply the explicit stereotype directly or replace the stereotype based on experience, as in previous research on stereotyping in the context of impression formation (e.g., Devine, 1989; Fiske & Neuberg, 1990); here, we found that stereotypes influence the process of learning about group members—a novel pattern representing the transmission of explicit stereotype knowledge to one's personal choice preference. Moreover, this transmission was evident despite participants' explicit goal to earn money based on accuracy, limited subjective

awareness of group members' actual sharing rates in Study 3, and disagreement with the messenger in Study 5.

Next, we asked whether these group choice preferences—once formed in response to stereotype exposure—could spread to others who witness these interactions, via *social learning* (Olsson, Knapska, & Lindström, 2020). Indeed, we found that stereotype-induced preferences in participants' choice behavior were transmitted to unwitting observers who, after viewing this behavior with no knowledge of group differences or intentions to discriminate, expressed this bias in their own choices. Together, these studies reveal a process through which exposure to social stereotypes may be internalized as one's personal group-based preference and transmitted to unsuspecting others.

This research introduces a model of intergroup bias that describes how stereotype knowledge, regardless of one's personal endorsement, can induce prejudice formation in intergroup interactions. This model builds on prior work in which stereotype knowledge, prejudiced attitudes, and direct experience have separate or competing roles in intergroup judgments and impressions (e.g., Amodio & Devine, 2006; Devine, 1989; Fiske & Neuberg, 1990; Kunda & Spencer, 2003), and it extends their purview to social-interactive contexts and choice behavior. Moreover, by describing a general mechanism through which societal knowledge can influence individual-level learning, this model contributes a novel framework for the interplay of societal- and individual-level forms of bias.

This work is notable for its use of computational modeling to examine stereotype function. Here, we adapted models of rule-based priors on reinforcement learning (e.g., Doll et al., 2009; Fareri, Chang, & Delgado, 2015) to address the effect of stereotype knowledge on interactive learning (Amodio & Cikara, 2021). By formalizing and comparing competing models, we found strong direct support for the hypothesized *biased learning model*, whereby stereotypes operated as priors and differentially affected learning from group members. This work illustrates how computational modeling may be used fruitfully to investigate mechanisms of social cognition and their interplay with features of society (Bai et al., 2022; FeldmanHall & Nassar, 2021; Kobayashi, Kable, Hsu, & Jenkins, 2022; Zhou et al., 2022).

More broadly, our findings show that messages promoting societal stereotypes are more than mere words; exposure to a biased political message can shape one's subsequent experiences with members of the group, perhaps without one's knowledge, in a way that confirms the message and spreads it to others. This process—whereby societal stereotypes are transmitted to personal group preferences—may also help to explain how systemic biases, such as institutional inequality, may be transmitted via stereotypes from social structures to the minds of individuals (Payne, Vuletich, & Lundberg, 2017). As society continues to grasp the impact of

polarizing sociopolitical rhetoric, from campaign ads to social media, our findings suggest that its influence may be more potent and far-reaching than previously thought. Yet, by illuminating the processes through which explicit societal messages may induce personal bias in the individual, these results promise new approaches to reducing their impact.

### References

- Allidina, S., & Cunningham, W. A. (2021). Avoidance begets avoidance: A computational account of negative stereotype persistence. *Journal of Experimental Psychology: General*.
- Amodio, D. M. (2019). Social Cognition 2.0: An interactive memory systems account. *Trends in Cognitive Sciences*, 23, 21-33.
- Amodio, D. M., & Cikara, M. (2021). The social neuroscience of prejudice. *Annual Review of Psychology*, 72, 439-469.
- Amodio, D. M. & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology*, 91, 652-661.
- Bai, X., Fiske, S. T., & Griffiths, T. L. (2022). Globally inaccurate stereotypes can result from locally adaptive exploration. *Psychological Science*, 33, 671-684.
- Behrens, T. E., Hunt, L. T., & Rushworth, M. F. (2009). The computation of social behavior. *science*, 324, 1160-1164.
- Darley, J.M., & Gross, P.H. (1983). A hypothesis confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44, 20-33.
- Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, 441, 876-879.
- Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature neuroscience*, 8, 1611-1618.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of personality and social psychology*, 56, 5-18.
- Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton trilogy revisited. *Journal of personality and social psychology*, 11, 1139-1150.
- Doll, B. B., Jacobs, W. J., Sanfey, A. G., & Frank, M. J. (2009). Instructional control of reinforcement learning: a behavioral and neurocomputational investigation. *Brain research*, 1299, 74-94.
- Fareri, D.S., Chang, L.J., Delgado, M.R. (2015). Computational substrates of social value in interpersonal collaboration. *Journal of Neuroscience*, 32, 9045-9052.
- FeldmanHall, O., & Nassar, M. R. (2021). The computational challenge of social learning. *Trends in Cognitive Sciences*, 25, 1045-1057.
- Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 2, pp. 357–411). Boston: McGraw-Hill.
- Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In *Advances in experimental social psychology* (Vol. 23, pp. 1-74). Academic Press.

- Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science*, 306(5703), 1940-1943.
- Gilbert, S. J., Swencionis, J. K., & Amodio, D. M. (2012). Evaluative vs. trait representation in intergroup social judgments: Distinct roles of anterior temporal lobe and prefrontal cortex. *Neuropsychologia*, 50, 3600-3611.
- Hackel, L. M., Kogon, D., Amodio, D. M., & Wood, W. (2022). Group value learned through interactions with members: A reinforcement learning account. *Journal of Experimental Social Psychology*, 99, 104267.
- Hein, G., Engelmann, J. B., Vollberg, M. C., & Tobler, P. N. (2016). How learning shapes the empathic brain. *Proceedings of the National Academy of Sciences*, 113, 80-85.
- Hilton, J. L., & Von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, 47, 237-271.
- Kawakami, K., Amodio, D. M., & Hugenberg, K. (2017). Intergroup perception and cognition: An integrative framework for understanding the causes and consequences of social categorization. *Advances in Experimental Social Psychology*, 55, 1-80.
- Kobayashi, K., Kable, J.W., Hsu, M. & Jenkins, A.C. (2022). Neural representations of others' traits predict social decisions. *Proceedings of the National Academy of Sciences, USA*, 119(22), e2116944119.
- Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological bulletin*, 129, 522.
- Kurdi, B., Gershman, S. J., & Banaji, M. R. (2019). Model-free and model-based learning processes in the updating of explicit and implicit evaluations. *Proceedings of the National Academy of Sciences*, 116(13), 6035-6044.
- Lindström, B., Selbing, I., Molapour, T., & Olsson, A. (2014). Racial bias shapes social reinforcement learning. *Psychological Science*, 25(3), 711-719.
- Olsson, A., Knapska, E., & Lindström, B. (2020). The neural and computational systems of social learning. *Nature Reviews Neuroscience*, 1-16.
- Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, 28(4), 233-248.
- Rothbart, M. (1981). Memory processes and social beliefs. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 145-181). Hillsdale, NJ: Erlbaum.
- Shohamy, D., & Wagner, A. D. (2008). Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events.
- Zhou, Y., Lindström, B. R., Soutschek, A., Kang, P., Tobler, P. N., & Hein, G. (2022). *Learning from ingroup experiences changes intergroup impressions*. bioRxiv 2021.11.02.466926.

## Methods and Supplemental Information for

Stillerman, B. S., Lindström, B. R., Schultner, D. T., Hackel, L. M., Hagen, D. R., Jostmann, N. B., & Amodio, D. M. *Transmission of societal stereotypes to individual preferences through social reinforcement learning*

## Table of Contents

Study 1 .....	17
Method .....	17
Results.....	20
Study 2 .....	22
Method .....	22
Results.....	23
Study 3 .....	23
Method .....	23
Results.....	25
Study 4 .....	28
Method .....	28
Results.....	30
Study 5 .....	32
Method .....	32
Results.....	34
Study 6 .....	36
Method .....	36
Results.....	38
Computational modeling .....	41
Supplementary References.....	51



## Study 1

*Overview.* In Study 1, we tested whether explicit descriptions of groups would bias participants' choices of who to interact with and win money from. Participants read descriptions of two fictional groups and then played an economic game with ostensible players from those groups. Group membership was, on average, not associated with reward probability and thus not beneficial cue for choice performance.

### Method

#### *Participants*

Sixty-nine students at University of Amsterdam received course credit for their participation as well as a performance-based monetary bonus, ranging from \$1.30 – \$1.70. In this and subsequent reported studies, we excluded participants who failed to reach a learning criterion of 50% accuracy for 30%-70% player pairs during the test phase (i.e., A-B and G-H; see below for details of test phase procedure). In Study 1, this exclusion criterion yielded a final sample size of  $N = 61$  (45 women, 16 men;  $M_{\text{age}} = 21.56$  years,  $SD_{\text{age}} = 5.20$  years).

Ethics approval was obtained from the human subjects institutional review board at the University of Amsterdam.

#### *Procedure*

*Introduction and manipulation.* Upon arrival to the lab and following informed consent, participants learned that they would play a money sharing game with players from two social-geographical groups. Before beginning the task, participants were given the following descriptions of these groups (with descriptions of Group A and B counterbalanced across participants):

“In the main task you will play an interactive money-sharing game with people from two different groups who come from different places. For the purpose of this study, we will refer to these groups as Group A and Group B, and their members will be represented by avatars. Members of Group A live in a more affluent society, where crime is low and most people have good jobs. People from Group A are often perceived to be trustworthy, honest, and generous to others, and they are proud of their success. Group B, by comparison, lives in a society that is economically poor, with a high rate of unemployment and serious crimes such as robbery, assault, and murder. People from Group B are often perceived to be hostile, untrustworthy, and dishonest.”

These descriptions were based on common societal stereotypes of White and Black Americans, respectively (Devine & Elliot, 1995), and which also correspond to common stereotypes to White (native) and Moroccan Dutch immigrants.

This stereotype information was followed by a note that, despite these generalizations, there is individual variability, and that the participant should pay attention to individual players' behavior:

“So, as you see, these groups are different in many ways. However, individuals within each group vary, too. You will need to learn about these people as you engage in repeated interactions in the task.”

Participants were then shown avatars representing players from each group, with color cues (blue vs. green clothing, darker vs. lighter hair) signaling group membership (all other features were matched between groups). Participants were assigned, in counterbalanced fashion, to view either all female or all male-appearing avatars (Figure S1), to control for potential target gender effects. Participants were instructed that these players had participated in a previous experiment in which they decided how many points (redeemable for a monetary bonus) to share. Participants were further told that different players shared different amounts, and they should learn who shared more often to win the most points.

*Categorization task.* To ensure that participants learned the group identity of each player, they completed a categorization task embedded in a standard 7-block implicit association test (IAT). The first and fifth blocks of this IAT required simple categorizations of player to their group category, with accuracy feedback. This IAT was repeated at the very end of the task. Although not the focus of these studies, Time 1 IAT data indicated that explicit group descriptions alone created a significant IAT effect, with preference expressed toward Group A, similar to much prior research (e.g., Gregg et al., 2006) showing that IAT scores can be driven by a variety of influences including novel explicit group beliefs. The same pattern of Group A preference was observed at Time 2. We did not analyze IAT scores further, and the measure was dropped from all subsequent studies.

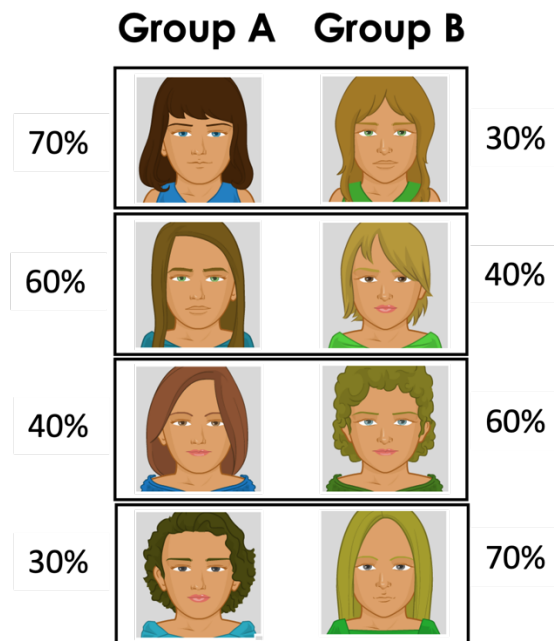


Figure S1. Sample avatars and schematic of reward probabilities.

*Learning task.* Next, participants completed the main learning task, which included a training phase of 160 trials and a test phase of 96 trials. In the training phase, participants

always chose between two targets—one from each group—with complementary fixed reward probabilities (e.g., player pairs A-B, C-D, E-F, and G-H, see Figure S1). On each trial, a face pair was shown for a maximum of 2 s, during which time a response was required. Reward feedback (+1 or 0 points) appeared immediately following choice, and points were converted to a cash bonus at the task conclusion. The cover story was that this feedback was derived from past participants' actual choices of how often to share points. Crucially, although the reward feedback varied by individuals within groups, average reward rate for both groups was equated. Player gender and group color cue (blue or green) were counterbalanced across participants, and player identity was randomized such that individual players were assigned to random reward rates for a given participant. To win as much money as possible, participants were motivated to learn which target players tended to reward more often than others.

Next, in the test phase, in order to obtain a readout of learned reward values, participants chose between all combinations of targets from different groups (e.g., A-B, A-D, A-F, A-H, C-B, etc.), always with one Group A member and one Group B member. Each pair was shown for a maximum of 2 s, during which time a response was required, followed by a 1000 ms intertrial interval. Feedback was not given, to prevent further learning, but participants were told that correct choices would still be rewarded and paid out in the bonus at the end of the task.

## Results

Primary analysis focused on test phase data and involved two approaches: multilevel regression and computational modeling. We detail the regression approach below and, for all studies, report computational modeling results in the section “Computational modeling”.

Multilevel regression was used to test effects of (a) players' actual reward rate and (b) group membership on choice. Trials in which choices were made faster than 200 ms or slower than 2000 ms were excluded from analysis. Participants' trial-level choice data were submitted to a general linear mixed model predicting the likelihood that participants chose a given target player, nested by participant, with a logit link function. The primary model included by-participant random intercepts and the following predictors as fixed effects: players' actual reward rate, players' group membership, and their interaction. For completeness, we also report models with by-participant random slopes for the fixed effects.

Results indicated a significant effect of player's relative reward rate on choice, demonstrating learning of player reward rates,  $B = 2.68$ ,  $SE = 0.19$ , Wald  $z = 14.43$ ,  $p < .001$ . An examination of raw choice behavior revealed a relatively accurate mapping between participants' choices and the actual reward contingencies. Importantly, the effect of group membership on choice was also significant, such that participants more likely to choose Group A members over Group B members,  $B = 0.52$ ,  $SE = 0.06$ , Wald  $z = 9.33$ ,  $p < .001$ . Indeed, when faced with two equally rewarding players, participants chose the Group A member 25% more often. The Reward Rate  $\times$  Group interaction was not significant,  $B = -0.003$ ,  $SE = 0.26$ , Wald  $z = -0.01$ ,  $p = .992$ . The pattern was qualitatively identical in the random slopes model (Reward rates:  $B = 3.27$ ,  $SE = 0.36$ , Wald  $z = 9.08$ ,  $p < .001$ , Group bias:  $B = 0.65$ ,  $SE = 0.23$ , Wald  $z = 2.79$ ,  $p = .005$ )

To corroborate these findings, training phase data were submitted to the same general linear mixed model. Results replicated those of the test phase data, with evidence of accurate learning of player reward rates,  $B = 1.77$ ,  $SE = 0.10$ , Wald  $z = 18.43$ ,  $p < .001$ , as well as a bias to

choose Group A,  $B = 0.29$ ,  $SE = 0.04$ , Wald  $z = 6.95$ ,  $p < .001$ . The reward rate by group interaction was marginally significant,  $B = -0.25$ ,  $SE = 0.13$ , Wald  $z = -1.83$ ,  $p = .068$ .

## Study 2

*Overview.* In Study 2, we sought to replicate the findings of Study 1 in a new sample. The procedure was identical to that of Study 1, except that the initial categorization task and IATs were dropped and it was conducted online rather than in the lab.

Ethics approval was obtained from the human subjects institutional review board at New York University.

## Method

*Participants.* Participants were 78 Amazon Mechanical Turk (MTurk) workers (demographics unavailable due to technical error) who received \$2.00 for their participation as well as a performance-based monetary bonus, ranging from approximately \$0.30 – \$0.40, derived from points earned during the task with a conversion rate of 2 points per cent. Participants who failed to reach a learning criterion of 50% accuracy for choices between 30% vs. 70% reward player pairs during the test phase (i.e., A-B and G-H;  $N = 16$ ) were excluded. After exclusions, Study 2 had a final sample size of  $N = 62$ .

*Procedure.* Participants completed an online learning task, nearly identical to the one described in Study 1. Besides a slightly different look and feel for the online version, the only difference was that the groups described with positive and negative stereotypes were counterbalanced (e.g., whether Group A or Group B was described as the good group). For ease

of reporting and visualizing the analysis, we refer to the group described with positive and negative stereotypes as “Group A” and “Group B,” respectively, in the results.

## Results

Our analytical approach followed that of Study 1, with a focus on test phase choice data. Results indicated significant effects for players’ actual reward rate, demonstrating strong learning,  $B = 2.55$ ,  $SE = 0.19$ , Wald  $z = 13.51$ ,  $p < .001$ , and for group membership, such that participants strongly preferred Group A members independent of actual reward rates,  $B = 0.79$ ,  $SE = 0.06$ , Wald  $z = 13.86$ ,  $p < .001$ . The Reward Rate x Group Membership interaction was not significant,  $B = -0.35$ ,  $SE = 0.26$ , Wald  $z = -1.32$ ,  $p = .187$ . As in Study 1, the qualitative pattern was identical in the random slopes model Reward rates:  $B = 3.03$ ,  $SE = 0.39$ , Wald  $z = 7.7$ ,  $p < .001$ , Group bias:  $B = 1.02$ ,  $SE = 0.37$ , Wald  $z = 2.8$ ,  $p = .005$ )

As in Study 1, to corroborate these findings, we submitted the training phase data to the same general linear mixed model. The results replicated those of the test phase data, with significant effects of actual reward rate,  $B = 1.40$ ,  $SE = 0.10$ , Wald  $z = 14.66$ ,  $p < .001$ , and of group membership, evidencing a preference for Group A members,  $B = 0.45$ ,  $SE = 0.04$ , Wald  $z = 10.59$ ,  $p < .001$ . The interaction was not significant,  $B = 0.17$ ,  $SE = 0.14$ , Wald  $z = 1.24$ ,  $p = .214$ .

## Study 3

*Overview.* In Study 3, we extended the procedure used in Studies 1 and 2 to include two additional post-learning-task measures: explicit beliefs of player reward rates and a trust game.

## Method

*Participants.* Participants were 158 Amazon Mechanical Turk (MTurk) workers who received \$2.00 for their participation as well as a performance-based monetary bonus, ranging

from approximately \$0.30 – \$0.40, derived from points earned during the task with a conversion rate of 2 points per cent. We excluded participants who responded without variation for either of the post-learning-task measures ( $N = 18$ ) and participants who failed to reach a learning criterion of 50% accuracy for 30%-70% player pairs during the test phase ( $N = 47$ ). One participant was excluded due to a technical error resulting in invalid post-learning-task measures. Our exclusion of trials with invalid reaction times resulted in 5 participants being excluded altogether. These exclusions resulted in a final sample size of  $N = 87$  (44 men, 36 women, 7 unreported;  $M_{age} = 34.9$  years,  $SD_{age} = 10.0$  years).

Ethics approval was obtained from the human subjects institutional review board at New York University.

*Procedure.* After reading the group instructions, participants completed a categorization task to reinforce the group membership of target players. Unlike the categorization task used in Study 1, which was embedded within an IAT, Study 3 used a stand-alone task that included the classification of both group member faces and trait terms that had been conveyed in the group description manipulation. Hence, participants were presented with pictures of players and stereotype words associated with the group descriptions (e.g., “wealthy,” “uneducated,” “trustworthy”) and classified each according to group label. They then completed the learning task, as in Study 2; however, in Study 3, Group A always associated with the positive stereotype description and Group B was associated with the negative stereotype description. This decision was made because A and B are often associated with better and worse options, respectively, and this could contribute to noise or confusion with the manipulation.



After the learning task, participants completed a subjective reward measure, in which they were asked, for each player in randomized order, “How many times out of a hundred would this player share with you?” For each player, participants typed their estimate of the player’s sharing rate, from 1 to 100, in a text box. This form of response was designed to assess declarative semantic knowledge, which might be expressed independent of any striatally-based instrumental tendencies that could influence responses on a slider-type scale (Knowlton et al. 1996).

Finally, participants played a single-shot trust game with each target player. They were told they had a 20-point pool and they could choose how much to invest in each player as a trustee. The trustee’s point amount would then be tripled and they could share any amount back to the participant. For each player, the participant selected a number of points to share, from 0 to 20, with options at 2-point intervals.

## Results

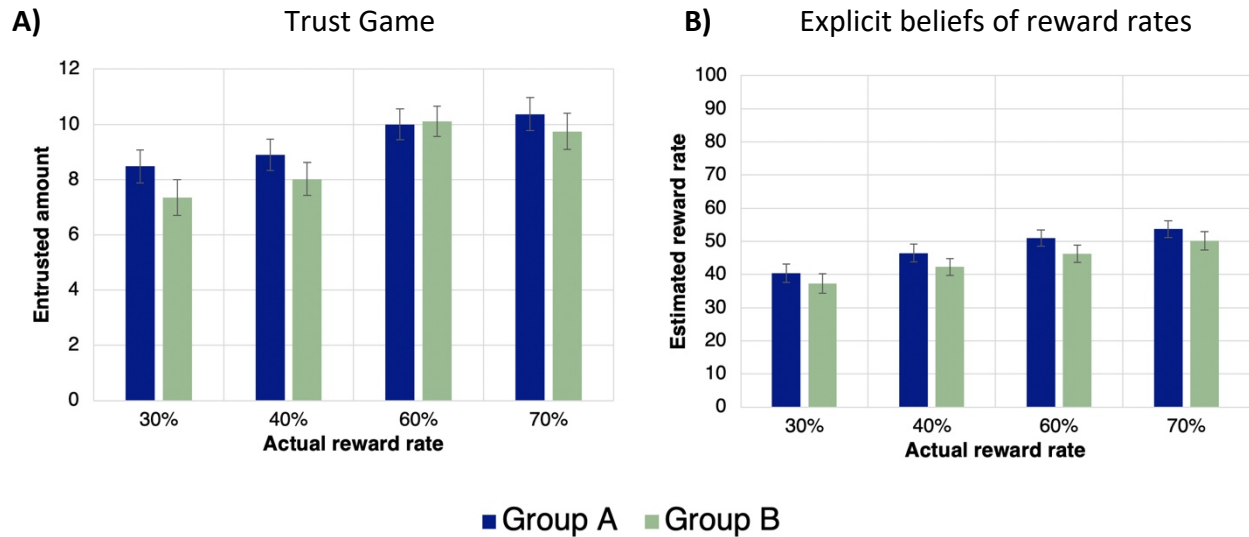
*Choice behavior.* Our analytical approach followed that of Studies 1 and 2. Multilevel regression predicting player choice again produced significant effects of players’ actual reward rate,  $B = 1.76$ ,  $SE = 0.16$ , Wald  $z = 11.15$ ,  $p < .001$ , as well as group membership, with participants preferring Group A members,  $B = 0.47$ ,  $SE = 0.05$ , Wald  $z = 9.58$ ,  $p < .001$ . Again, the interaction was not significant,  $B = 0.01$ ,  $SE = 0.23$  Wald  $z = 0.05$ ,  $p = .964$ . Again, the random slopes model gave the same pattern (Reward rates:  $B = 2.01$ ,  $SE = 0.31$ , Wald  $z = 6.5$ ,  $p < .001$ , Group bias:  $B = 0.52$ ,  $SE = 0.19$ , Wald  $z = 2.74$ ,  $p = .006$ )

As in Studies 1 and 2, analysis of training phase data produced the same pattern: significant effects of players’ reward rate,  $B = 0.93$ ,  $SE = 0.08$ , Wald  $z = 11.04$ ,  $p < .001$ , and of

group membership,  $B = 0.46$ ,  $SE = 0.04$ , Wald  $z = 12.09$ ,  $p < .001$ , on choice. The interaction was not significant,  $B = 0.04$ ,  $SE = 0.12$ , Wald  $z = 0.40$ ,  $p = .691$ .

*Trust game behavior.* Participants' trust game investments were submitted to a linear regression, with players' (trustee) true reward rate and player (trustee) group as predictors. Participants' investments were significantly predicted by players' actual reward rates,  $B = 5.87$ ,  $SE = 1.34$ ,  $t(693) = 4.38$ ,  $p < .001$ , 95% CI [3.24, 8.50], reflecting that reward learning translated to an expression of trust. However, the effect of group membership was not significant,  $B = 0.63$ ,  $SE = 0.42$ ,  $t(693) = 1.48$ ,  $p = .140$ , 95% CI [-0.19, 1.45].

*Subjective rewards.* Participants' subjective reward estimates were submitted to a linear regression, with actual player reward rate and player group as predictors. Subjective estimates were significantly predicted by players' actual reward rates,  $B = 30.59$ ,  $SE = 6.01$ ,  $t(683) = 5.09$ ,  $p < .001$ , indicating participants had some knowledge of the reward contingencies. Subjective reward rates were also predicted by group membership,  $B = 3.88$ ,  $SE = 1.90$ ,  $t(683) = 2.04$ ,  $p = .042$ , suggesting a weak effect of group bias on subjective reward in addition to the relatively strong effect effects observed on choice behavior.



*Figure 3.* (A) Number of points entrusted to players during the trust game. Study 3 participants entrusted more points to the more rewarding players and to Group A members. (B) Estimated reward rates for each player. Study 3 participants estimated higher reward rates for more rewarding players and for Group A members.

To test whether the effects of group membership on choice behavior could be explained by subjective rewards, or whether the group bias instrumental learning operated independently of subjective rewards, we conducted an analysis in which player reward rates, group membership, and subjective estimates were each included in the main multilevel model predicting test phase choice. We found that while subjective rewards predicted choices to a small extent,  $B = 0.03$ ,  $SE = 0.01$ , Wald  $z = 26.62$ ,  $p < .001$ , actual reward rates ( $B = 0.94$ ,  $SE = 0.11$ , Wald  $z = 5.44$ ,  $p < .001$ ) and player group ( $B = 0.21$ ,  $SE = 0.05$ , Wald  $z = 3.97$ ,  $p < .001$ ) remained strong predictors of choice behavior. This result suggests that subjective beliefs about group differences in reward did not fully account for the group effect expressed in behavior.

## Study 4

*Overview.* In Study 4, we asked whether the prejudice could be propagated through mere observation of biased choices without any knowledge of group descriptions or direct feedback from group members. In this study procedure, participants observed the training data of a past participant, viewing their choice and feedback but not being exposed to any group descriptions, and then completed their own test phase choices. Each Study 4 participant was yoked to real Study 2 participant (“demonstrator”) and observed the training phase behavior of the participant to whom they were yoked. To test whether participants formed a group bias based on this observational learning, Study 4 participants then made their own choices in the test phase. After the test phase of the learning task, participants reported their estimates of player reward rates in the same explicit belief measure as Study 3, followed by estimates of the choice behavior of the demonstrators that they observed. Study 4 was pre-registered at <https://aspredicted.org/blind.php?x=6zi6fz>.

## Method

*Participants.* Participants were recruited from the online NYU subject pool and received course credit for their participation as well as the chance to win a performance-based monetary bonus of \$15 to be awarded to the five top performers. Our stopping rule was to collect data until two Study 4 participants were yoked to each of the 62 demonstrators from Study 2 (for a total  $N$  of 124). We excluded participants who responded without variation in the post-learning task ( $N = 6$ ), participants who failed to reach 50% on either attention check measure ( $N = 13$ ; see Procedure), and participants who failed to reach a learning criterion of 50% accuracy for the 30%-70% pairs during the test phase (i.e., A-B and G-H;  $N = 33$ ). Because we could not know the

number of eventual exclusions during the period of data collection, data were collected from an additional 36 participants who met inclusion criteria but were ultimately not needed and thus excluded. These extra participants were excluded based on chronological order of completing the experiment, and their data were never analyzed. After all exclusions, this approach yielded the target final sample size of  $N = 124$  (82 women, 39 men, 3 unreported;  $M_{age} = 19.5$  years,  $SD_{age} = 1.35$  years).

Ethics approval was obtained from the human subjects institutional review board at New York University.

*Procedure.* Participants read instructions similar to Studies 1 – 3, which explained the nature of the sharing task and the fact that players represented two different groups, but they received no descriptions of the groups. They were also told that they were to learn about the target players by watching past “demonstrators” make decisions and receive feedback. Participants then completed a categorization task of the group membership of target players, which served to both reinforce group membership cues and provide an attention check on which to base participant exclusions. Unlike in Study 3, this categorization task did not include stereotype words; participants only categorized faces of group members. Participants with less than 50% accuracy on the classification task were excluded ( $N = 9$ ).

Next, participants observed the training phase, where, instead of making choices, participants observed the trials of the Study 2 demonstrator to whom they were yoked. Trials were presented in the same order and each yoked trial was animated in real time (using actual reaction times for each prior demonstrator choice) to show choices and subsequent reward feedback, identical to how direct learners viewed choices and feedback. Participants observed

the entirety of the yoked training phase, complete with a break between the two blocks, with the exception that trials were skipped if they had originally been excluded in Study 2 based on reaction time. To ensure participants paid attention, “catch” trials appeared after some trials, prompting participants to indicate what choice they had just observed on the previous trial. Twenty catch trials appeared in the observational training phase, occurring in a fixed, pseudorandom order. Participants with less than 50% accuracy on the catch trials were excluded ( $N = 4$ ).

Participants then completed the test phase, making their own choices, as in Studies 1 – 3. Next, participants reported their explicit beliefs about player reward rates, as in Study 3, typing their response in a box under displays of each player (“How many times out of a hundred would this player share with you?”). Finally, participants reported their estimates of demonstrators’ tendency to choose each player (“How many times out of a hundred did the Decider choose this player?”).

## Results

*Observational learning effects.* Our analytical approach followed that of Studies 1 – 3, with a focus on test phase choices. In this study, however, participants did not directly complete a training phase, but instead observed training phase behavior of Study 2 participants. As in the previous studies, in which learning occurred directly, multilevel regression indicated that observational learning produced a significant effect of actual reward rates,  $B = 1.61$ ,  $SE = 0.13$ , Wald  $z = 12.74$ ,  $p < .001$ , as well as a significant effect of group membership, with Study 4 participants preferring Group A players,  $B = 0.32$ ,  $SE = 0.04$ , Wald  $z = 8.03$ ,  $p < .001$ . The interaction was not significant,  $B = -0.24$ ,  $SE = 0.18$ , Wald  $z = -1.36$ ,  $p = .173$ .

As in the preceding experiments, the random effects analysis produced a qualitatively identical pattern (Reward rates:  $B = 1.78$ ,  $SE = 0.22$ , Wald  $z = 7.99$ ,  $p < .001$ , Group bias:  $B = 0.37$ ,  $SE = 0.19$ , Wald  $z = 2.0$ ,  $p = .045$ ).

It should be noted that this analysis deviated slightly from our preregistered plan, which was to yoke Study 4 participants to the total Study 2 sample ( $N = 78$ ), with two Study 4 participants yoked to each Study 2 participant in order to increase power. This pre-registration did not consider that some Study 2 participants would provide invalid or incomplete data. Hence, in order to obtain validity and rigor, Study 4 participants were yoked only to Study 2 participants included in the final Study 2 analysis. Nevertheless, results were nearly identical using this sample ( $N = 156$ ): test phase learning effect:  $B = 1.49$ ,  $SE = 0.11$ , Wald  $z = 13.36$ ,  $p < .001$ ; group effect:  $B = 0.39$ ,  $SE = 0.03$ , Wald  $z = 11.25$ ,  $p < .001$ ; interaction:  $B = -0.17$ ,  $SE = 0.18$ , Wald  $z = -1.13$ ,  $p = .258$ .

In addition to these analyses, we also estimated the direct transmission of bias by predicting the participants' preference for Group A from the demonstrator's Group A bias using multilevel regression ( $B = 0.28$ ,  $SE = 0.09$ , Wald  $z = 3.21$ ,  $p = .001$ ). This result indicates that the degree of demonstrator group preference was significantly correlated with the degree of observer group preference.

*Explicit beliefs.* In the reward estimation task, participants' estimates of targets' players' reward rates was not significantly associated with those players' actual reward rates,  $B = -2.48$ ,  $SE = 4.99$ ,  $t(989) = -0.50$ ,  $p = .619$ , indicating participants had very poor, if any, declarative knowledge of the reward contingencies. There was also no evidence of a group bias in estimations of reward rates,  $B = 0.79$ ,  $SE = 1.58$ ,  $t(107) = 0.50$ ,  $p = .615$ . Thus, the observational

learning of bias appeared to emerge in the absence of explicit beliefs or knowledge regarding player reward rates.

As in Study 3, to more directly test whether choice behaviors reflected a group bias in the absence of explicit beliefs, we tested the main regression with actual player reward rate, group membership, and explicit belief estimates as predictors. Results indicated a small-effect size association between explicit beliefs and choice behavior,  $B = 0.01$ ,  $SE = 0.001$ , Wald  $z = 20.22$ ,  $p < .001$ , significant effects remained for actual reward rates,  $B = 1.69$ ,  $SE = 0.13$ , Wald  $z = 13.09$ ,  $p < .001$ , and group membership,  $B = 0.26$ ,  $SE = 0.04$ , Wald  $z = 6.52$ ,  $p < .001$ .

An analysis of observers' estimate of demonstrator choices indicated that these did not significantly reflect the actual player reward rates,  $B = -0.66$ ,  $SE = 4.77$ ,  $t(989) = -0.14$ ,  $p = .890$ , or group membership,  $B = 1.05$ ,  $SE = 1.51$ ,  $t(989) = 0.70$ ,  $p = .486$ .

## Study 5

*Overview.* In Study 5, we asked whether the pattern of group membership bias on reward learning and choice would emerge even if the stereotypic group descriptions came from a politician with whom the participant strongly disagrees. The procedure was similar to that of Study 3, except that group descriptions were attributed to President Donald Trump and participants indicated their political orientation (Democrat or Republican) and agreement with Donald Trump. Study 5 was preregistered at <https://aspredicted.org/blind.php?x=fv9bd4>.

## Method

*Participants.* A total of 318 Amazon Mechanical Turk (MTurk) workers participated in the experiment in return for \$2.00 plus a performance-based monetary bonus, ranging from approximately \$0.30 – \$0.40, derived from points earned during the task with a conversion rate



of 2 points per cent. We excluded participants who responded without variation in the post-learning task ( $N = 7$ ) and participants who failed to reach a learning criterion of 50% accuracy for 30%-70% player pairs during the test phase (i.e., A-B and G-H;  $N = 65$ ). Our exclusion of trials with invalid reaction times resulted in 3 additional participant exclusions. Because we could not know the eventual number of exclusions during data collection, an additional 43 participants who would have met the inclusion criteria were collected but were ultimately excluded because the final sample goal had already been reached. These participants were excluded based on chronological order of completing the experiment. When they are included in the analysis, the results are nearly identical. After all exclusions, Study 5 had a final sample size of  $N = 200$  (138 Democrats, 62 Republicans).

Ethics approval was obtained from the human subjects institutional review board at New York University.

*Procedure.* Participants read similar instructions to Study 3, but the descriptions of the groups were described as comments from President Donald Trump. Participants learned they would complete a money sharing task with players who come from two different communities, and that President Trump described one group as “coming from a society that is economically poor, with a high rate of unemployment and serious crimes such as robbery, assault, and murder. He has described people from this group as hostile, untrustworthy, and dishonest.” Participants then learned the President Trump had not commented on the second group, from which other players came, and thus the second group in this study constituted a neutral control group. Participants then completed a categorization task, as in Study 3, to reinforce the group membership of target players. In this task, they were presented with pictures of players and

stereotype words associated with the group descriptions (e.g., “wealthy,” “uneducated,” “trustworthy”). They then completed the learning task, as in Study 3. After the learning task, participants completed the same explicit belief of reward rates task as in Study 3, which included training and test phases. Finally, they answered three questions about their political orientation: free response of their political views (not analyzed here), a forced choice between Democrat and Republican, and rating of the extent to which they agree with President Trump on 5-point Likert scale, ranging from “Never” to “Always.”

## Results

Multilevel regression analysis of test phase choice behavior indicated significant effects of actual player reward rate,  $B = 1.98$ ,  $SE = 0.10$ , Wald  $z = 19.40$ ,  $p < .001$ , as well as group membership, with a preference for Group A members over Group B members,  $B = 0.40$ ,  $SE = 0.03$ , Wald  $z = 12.78$ ,  $p < .001$ . As in previous studies, the interaction was not significant,  $B = -.03$ ,  $SE = 0.14$ , Wald  $z = -0.23$ ,  $p = .821$ . The random effects analysis produced a qualitatively identical pattern (Reward rates:  $B = 2.45$ ,  $SE = 0.22$ , Wald  $z = 11.15$ ,  $p < .001$ , Group bias:  $B = 0.57$ ,  $SE = 0.16$ , Wald  $z = 3.61$ ,  $p = .0003$ ).

Corroborating these findings, the same regression model applied to training phase data also indicate significant effects of actual player reward rate,  $B = 1.13$ ,  $SE = 0.05$ , Wald  $z = 20.98$ ,  $p < .001$ , as well as group membership,  $B = 0.38$ ,  $SE = 0.02$ , Wald  $z = 15.75$ ,  $p < .001$ . However, in this analysis, the Reward rate x Group membership interaction was significant,  $B = 0.18$ ,  $SE = 0.08$ , Wald  $z = 2.40$ ,  $p = .016$ , indicating the group effect on choices was greater when the given player was more rewarding than the other option.

Next, we tested whether the effect of Trump's group message biased choice behavior even among participants identified with the opposing Democratic party. Analysis of participants' post-task responses showed that those who identified as Democrats strongly disagreed with Donald Trump, in comparison to participants who identified as Republicans.

To examine the effect of political identity on group biases in learning, we first tested whether group effects on choice were moderated by political identity, including political identity with player reward rate and group membership in the regression model. In addition to significant effects for player reward rate,  $B = 1.98$ ,  $SE = 0.12$ , Wald  $z = 15.99$ ,  $p < .001$ , and group membership,  $B = 0.31$ ,  $SE = 0.04$ , Wald  $z = 8.15$ ,  $p < .001$ , as reported above, this analysis revealed a significant Group x Political identity interaction,  $B = 0.32$ ,  $SE = 0.07$ , Wald  $z = 4.58$ ,  $p < .001$ , whereby Republicans exhibited a stronger effect of group membership on choice behavior than Democrats. However, separate tests of the main regression model for Democrat and Republican participants indicated that participants from both parties showed a significant group bias effect (Democrats:  $B = 0.31$ ,  $SE = 0.04$ , Wald  $z = 8.14$ ,  $p < .001$ ; Republicans:  $B = 0.62$ ,  $SE = 0.06$ , Wald  $z = 10.83$ ,  $p < .001$ ). That is, although the group stereotype expressed by President Trump had a stronger impact on Republican participants' learning than on Democrats' learning, even Democratic participants were significantly influenced by Trump's message.

Analysis of explicit beliefs of reward rates indicated that these estimate were significantly associated with actual player reward rates,  $B = 40.79$ ,  $SE = 3.68$ ,  $t(1597) = 11.09$ ,  $p < .001$ , although these explicit estimates explained only 7.1% of the variance in actual rates. However, explicit beliefs were not significantly related to group membership,  $B = -1.13$ ,  $SE =$

1,16,  $t(1597) = -0.97$ ,  $p = .334$ , suggesting that participants may have been unaware of the group bias exhibited in their choice behavior. As in Studies 3 and 4, we included explicit beliefs as an additional predictor of learning task choices in the main multilevel model to test the relationship between explicit beliefs and choice behavior directly. We found that, although explicitly-estimated reward rates predicted choices to a small extent (i.e., with a small effect size),  $B = 0.03$ ,  $SE = 0.01$ , Wald  $z = 42.07$ ,  $p < .001$ , actual player reward rates,  $B = 0.90$ ,  $SE = 0.11$ , Wald  $z = 8.10$ ,  $p < .001$ , and group membership ( $B = 0.39$ ,  $SE = 0.03$ , Wald  $z = 11.47$ ,  $p < .001$ ) remained as stronger predictors of choice behavior. This indicates that participants' choice behavior, including their group bias, could not fully be accounted for by participants' explicit beliefs of player reward rates.

## Study 6

*Overview.* Study 6 tested whether the group bias biases acquired from a politician—even one that a learner objects to—can be propagated to others who merely observe the learner's behavior. To this end, Study 6 used an identical procedure to that of Study 4, but the underlying yoked data that participants observed was drawn from Study 5 demonstrators. As in Study 4, Study 6 participants viewed the training data of past participants but made their own choices in the test phase.

### Method

*Participants.* Participants were recruited from MTurk and were paid \$2.00 plus a performance-based monetary bonus, ranging from approximately \$0.30 – \$0.40, derived from points earned during the task with a conversion rate of 2 points per cent. Our stopping rule was to collect data until each Study 6 participant was yoked to a Study 5 demonstrator. Prior to

Study 6, eleven Study 5 demonstrators were removed because too many of their training phase trials included invalid (too fast) reaction times that precluded an adequate opportunity for observational learning in Study 6. This resulted in a target sample of  $N = 189$ . We continued data collection until we reached our target sample. We excluded Study 6 participants who responded without variation for either of the post-learning-task measures ( $N = 6$ ), failed either of two attention checks ( $N = 58$ ; see Procedure), or failed to reach a learning criterion of 50% accuracy for the 30%-70% pairs during the test phase (i.e., A-B and G-H;  $N = 53$ ). Because we could not know in advance how many participants would be excluded, an additional 72 participants who would have met the inclusion criteria were excluded because the Study 5 demonstrator to whom they were yoked had already been successfully matched with a Study 6 participant. These Study 6 participants were excluded based on chronological order of completing the experiment, and their inclusions results in a nearly identical pattern of results. After all exclusions, Study 6 had a final sample size of  $N = 189$ .

Ethics approval was obtained from the human subjects institutional review board at New York University.

*Procedure.* As in Study 4, participants were told that they were to learn about target players by watching past “demonstrators” make decisions and receive feedback. Prior to completing this learning task, participants completed the categorization task of the group membership of target players (as in Study 4). Participants with less than 50% accuracy on the classification task were excluded ( $N = 32$ ).

Next, participants began the learning task. First, they observed demonstrators’ completion of the training phase, where demonstrators were Study 5 participants to whom

Study 6 participants were yoked. Trials were presented in the order original completed by each demonstrator, and each yoked trial was animated in real time, using actual reaction times, to show choices and subsequent reward feedback. Participants observed the entirety of the yoked training phase, complete with a break between the two blocks, with the exception that trials were skipped if they had originally been excluded in Study 5 due to extreme reaction times. To ensure participants paid attention, “catch” trials appeared after some trials, prompting participants to indicate what choice they had observed on the previous trial. Twenty catch trials appeared in the observational training phase, occurring in a fixed, pseudorandom order. Participants with less than 50% accuracy on the catch trials were excluded ( $N = 24$ ). Participants then completed the test phase, making their own choices, as in the previous studies.

Next, participants provided their explicit beliefs of player reward rates, as in Study 4, where they typed their estimates of the reward rates for each player (“How many times out of a hundred would this player share with you?”) and their estimates of demonstrator choice behavior (“How many times out of a hundred did the Decider choose this player?”).

## Results

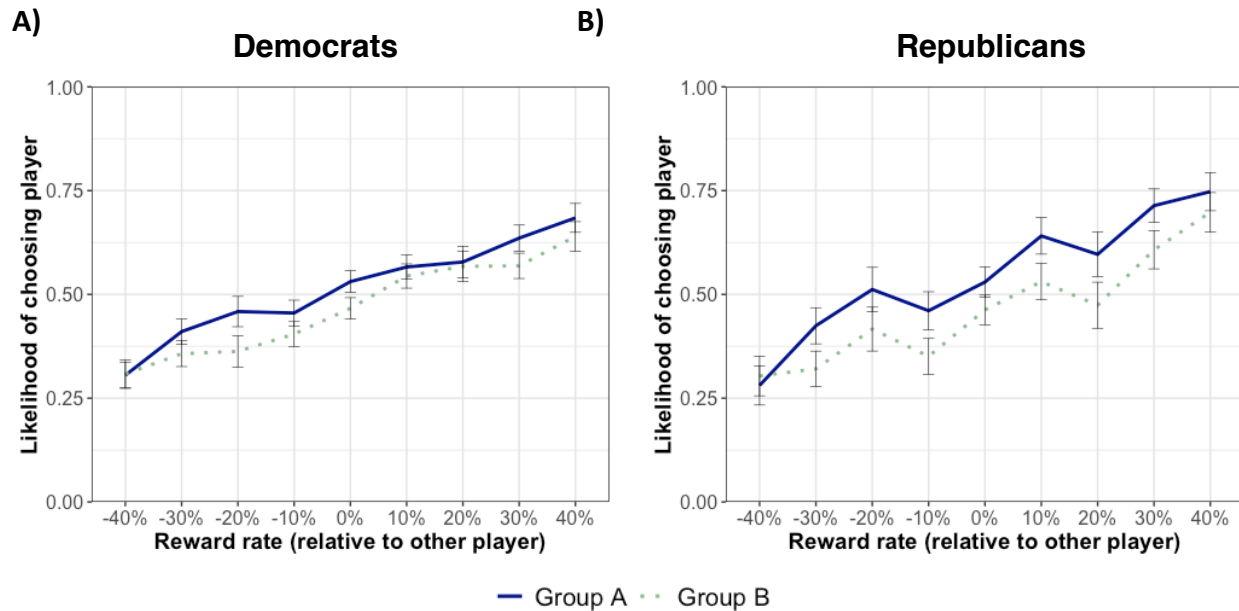
Multilevel regression analysis of choice behavior indicated significant effects of the actual player reward rates experienced by the original demonstrators ( $B = 1.82$ ,  $SE = 0.10$ , Wald  $z = 17.43$ ,  $p < .001$ ), and of players’ group memberships, such that Study 6 participants preferred members of the group not disparaged by Trump to demonstrators ( $B = 0.25$ ,  $SE = 0.03$ , Wald  $z = 7.98$ ,  $p < .001$ ). The interaction was not significant ( $B = 0.16$ ,  $SE = 0.15$  Wald  $z = 1.11$ ,  $p = .268$ ). In contrast, the random slopes model suggested that the average effect of group highly variable ( $B = 0.37$ ,  $SE = 0.24$ , Wald  $z = 1.53$ ,  $p = 0.13$ ), while the effect of reward

rate was highly significant ( $B = 2.74$ ,  $SE = 0.24$ , Wald  $z = 11.67$ ,  $p < .001$ ). However, as in Study 4, and regardless of model specification, there was a significant transmission of bias from demonstrator to participant ( $B = 0.28$ ,  $SE = 0.11$ , Wald  $z = 2.62$ ,  $p = .009$ ). Thus, even though Study 6 participants had no information about group descriptions given to Study 5 demonstrators, they nevertheless showed this bias in their own choice preferences. Moreover, the more biased demonstrators were, the stronger was this biased preference.

Next, we tested whether the degree of group bias acquired by observers depended on the political identity of demonstrators. To this end, we included demonstrator political identity as a predictor along player reward rate and group membership in the main multilevel regression. This analysis revealed that, in addition to main effects reward rate,  $B = 1.70$ ,  $SE = 0.13$ , Wald  $z = 13.43$ ,  $p < .001$ , and group membership,  $B = 0.21$ ,  $SE = 0.04$ , Wald  $z = 5.26$ ,  $p < .001$ , the effect of group membership on choice behavior was moderated by demonstrator political identity,  $B = 0.15$ ,  $SE = 0.07$ , Wald  $z = 2.20$ ,  $p = .028$ , such that observers of Republican demonstrators expressed a more pronounced group bias than observers of Democrat demonstrators (see Figure S2). Despite this difference, separate tests showed that observers acquired significant group bias regardless of whether they observed the behavior of a Republican or Democrat demonstrator (Republican demonstrator:  $B = 0.35$ ,  $SE = 0.05$ , Wald  $z = 6.36$ ,  $p < .001$ ; Democrat demonstrator:  $B = 0.21$ ,  $SE = 0.04$ , Wald  $z = 5.28$ ,  $p < .001$ ).

In the reward estimation task, participants' estimates of targets players' reward rates were not significantly associated with those players' actual reward rates,  $B = -0.78$ ,  $SE = 3.83$ ,  $t(1509) = -0.20$ ,  $p = .839$ , indicating participants did not have accurate explicit knowledge of player's actual reward rates, despite evidence of having learned these reward rates in behavior.

The effect of group membership on explicit beliefs of reward rates was not significant,  $B = -2.01$ ,  $SE = 1.21$ ,  $t(1509) = -1.66$ ,  $p = .097$ .



*Figure S2.* Test phase choice behavior for Study 6 (Panels A and B depict observers yoked to Democrat and Republican demonstrators, respectively). Participants observing both political parties demonstrated successful learning of reward contingencies and a group bias, but this was larger for observers of Republicans. Reward rate was the actual reward rate of a given player minus the actual reward rate of the other player option in a trial. Error bars represent standard error.

To test the degree to which the group bias in observers' choice behavior could be explained by their explicit beliefs about player reward rates, we included explicit beliefs as an additional predictor of choice, along with actual player reward rate and group membership, in the main multilevel model. Although explicitly-estimated reward rates predicted choices to a small extent, with a small effect size ( $B = 0.02$ ,  $SE = 0.001$ , Wald  $z = 31.26$ ,  $p < .001$ ), the influence of actual reward rates ( $B = 1.97$ ,  $SE = 0.11$ , Wald  $z = 18.20$ ,  $p < .001$ ) and group



membership ( $B = 0.19$ ,  $SE = 0.03$ , Wald  $z = 5.70$ ,  $p < .001$ ) remained as strong predictors. This indicates that participants' choice behavior, including their group bias, could not fully be accounted for by differences in their explicit beliefs about reward rates.

An analysis of participants' estimates of demonstrators choice preferences indicate that these estimates were not predicted by actual player reward rates,  $B = -3.18$ ,  $SE = 3.72$ ,  $t(1509) = -0.85$ ,  $p = .395$ . However, estimates of demonstrator choice did differ based on group,  $B = 4.94$ ,  $SE = 1.18$ ,  $t(1509) = 4.18$ ,  $p < .001$ , such that observers (correctly) estimated that that demonstrators chose Group A players more often—a pattern that was not found among observational learners in Study 4.

### Computational modeling

Our computational modeling analysis evaluated different hypotheses about the mechanisms underlying the group bias observed in the experiments. To this end, we adapted reinforcement learning (RL) and Bayesian learning models previously developed for understanding the influence on verbal instruction on learning (Doll, Jacobs, Sanfey, & Frank, 2009).

*Reinforcement learning.* The basis for all reinforcement (RL) models was the standard Q-learning (or Rescorla-Wagner) learning rule:

$$Q_i^{t+1} = Q_i^t + \alpha(R^t - Q_i^t) \quad [1]$$

where  $Q_i$  is the action value of selecting option  $i$  in trial  $t$ ,  $R$  is the reinforcement [no reward = 0, reward = 1] received in trial  $t$ , and  $\alpha$  ( $0 \leq \alpha \leq 1$ ) is a learning rate parameter, which

determines how much the difference between the received and the predicted reinforcement (the prediction error) affects subsequent value estimates (Rescorla & Wagner, 1972).

In all RL models, the Q-values were transformed into decision probabilities using a standard Softmax function

$$P_i = \frac{e^{Q_i/\beta}}{\sum_{j=1}^2 e^{Q_j/\beta}} \quad [2]$$

where  $\beta$  ( $0.01 < \beta \leq 100$ ) is the temperature parameter that determines the sensitivity of choices to the difference in Q-values. Very low values of  $\beta$  results in selecting the action with higher Q-value with probability  $\sim 1$ , while high values of  $\beta$  result in explorative choices that are insensitive to the difference in Q-values. Together, equations 1-2 gives an unbiased standard learning model (model 1).

We considered two main mechanisms for group-based bias in RL. First, the semantic information provided in the manipulated group descriptions could result in different *priors*, or initial expectancies, about the value of selecting each group at the outset of the training phase. We implemented this by estimating a prior parameter,  $P$  ( $-100 \leq P \leq 100$ ), which determined the initial Q value for the groups ( $Q_{Good}^{t=0} = \text{prior}$ ,  $Q_{Bad}^{t=0} = -\text{prior}$ ). In models without this parameter, the initial Q-values were set to be equal ( $Q_{Good}^{t=0} = Q_{Bad}^{t=0} = 0.5$ ). Non-zero values of the  $P$  parameter bias initial choices of the group with  $P > 0$ . We implemented a model with priors but no reward learning in the *bias prior* model (model 2).

In addition to models with such symmetric priors, we evaluated models with separate priors, in which initial expectancies for either group were allowed to vary independently,

introducing an extra parameter but allowing for increased flexibility. Models with symmetric priors provided a better fit to the data and their output will be reported in later sections.

It should be noted that, if the model allows for reward learning, the influence of the  $P$  parameter decreases exponentially across training trials. In other words, experiential learning can rapidly counteract the initial expectancies. We evaluated this in the *bias prior RL* model (model 3).

Second, the *learning rate*,  $\alpha$ , might differ between groups, so that participants update Q-values more (or less) rapidly from interacting with one group than the other (eq. 1). To evaluate biased updating, we either estimated  $\alpha$  by group (2  $\alpha$ ), or by both group and sign of the prediction error (4  $\alpha$ ), based on classic social psychological theories that relate prejudice to differential attention to groups (i.e., *ingroup favoritism*, Brewer, 1999; and *outgroup homogeneity*, Park & Rothbart, 1982) and differential processing of positive and negative behaviors of ingroup vs. outgroup members (i.e., the *ultimate attribution error*, Pettigrew, 1979)). Following earlier research (ref), we also evaluated a *confirmation bias* model (with confirmation bias parameter  $C$ ,  $1 \leq C \leq 10$ ), which amplifies gains and reduces losses for interactions with the “good” group (i.e.,  $C a_{positive} \delta_+$ ,  $(\frac{a_{negative}}{C}) \delta_-$ ), and vice versa for the bad group (i.e.,  $(\frac{a_{positive}}{C}) \delta_+$ ,  $C a_{negative} \delta_-$ ) (see Table S1 for overview).

*Bayesian learning.* We also tested how Bayesian learning models accounted for the data. The main motivation for this approach is that Bayesian priors can have a stronger, more long-lasting effect on behavior than in the RL framework we describe above (where the prior is just the initial Q-value). We used standard Bayesian beta-binomial learning models (Doll et al.,

2009), which explicitly estimate the probability of reward for selecting each group, given a beta distributed prior with hyperparameters  $\alpha$  and  $\beta$  (both initialized to 1 for each stimulus  $i$ ). The model learned by updating  $\alpha$  and  $\beta$  (for each stimulus) by adding the running count of reward and no-reward feedback (separately for each stimulus  $i$ ). Given a beta prior, this amounts to calculating the posterior distribution for each stimulus using Bayes rule:

$$\alpha_i^{t+1} = \alpha_i^t + pos \quad [3]$$

$$\beta_i^{t+1} = \beta_i^t + neg \quad [4]$$

where  $pos = 1$  after reward feedback, and 0 after no-reward feedback, and vice-versa for  $neg$ .

In addition, the running counts are decayed multiplicatively on each trial by a free parameter  $\gamma$  ( $0 \leq \gamma \leq 1$ ), which allows the model to forget potentially outdated information (Pettigrew, 1979). Choices were probabilistically taken (following a Softmax function, eq. 2) by comparing the modes of the posterior distributions:

$$mode_i = \frac{\alpha_i - 1}{\alpha_i + \beta_i - 2} \quad [5]$$

We evaluated two versions of this model. In the first version (model 9), the initial  $\alpha$  parameter for Group A was estimated. In this model formulation, both the mode and the precision of the prior is affected by  $\alpha$ . More evidence is required to counteract a precise prior. If  $\alpha_{Good}$  is higher than a  $\alpha_{Bad}$ , the model is biased to select group A. The second model (model 10) incorporated an additional parameter  $w$  ( $1 \leq w \leq 100$ ), which modulated the feedback in a

manner congruent with the semantic information (i.e., a confirmation bias). For stimuli from Group A, this gives

$$\alpha_i^{t+1} = \alpha_i^t + w_{pos} \quad [6]$$

$$\beta_i^{t+1} = \beta_i^t + \frac{1}{w} neg \quad [7]$$

In other words, the model learns faster from positive outcomes and slower from negative outcomes. For stimuli from Group B, the effect of  $w$  was the inverse (i.e., faster learning from negative outcomes and slower learning from positive outcomes).

Model #	Conceptual label	parameters	# parameters	Type
1	<i>Unbiased learning</i>	$\alpha, \beta$	2	RL
2	<i>Stereotype-only</i>	$\beta, P$	2	RL
3	<i>Stereotype-individuation</i>	$\alpha, \beta, P$	3	RL
4	<i>Group-learning</i>	$\alpha_{Good}, \alpha_{Bad}, \beta$	3	RL
5	<i>Biased learning</i>	$\alpha_{Good}, \alpha_{Bad}, P, \beta$	4	RL
6	/	$\alpha_{Good}, \alpha_{Bad}, C, \beta$	4	RL
7	<i>gain/loss group-learning</i>	$\alpha_{Good+}, \alpha_{Good-}, \alpha_{Bad+}, \alpha_{Bad-}, \beta$	5	RL
8	<i>Stereotype-gain/loss group learning</i>	$\alpha_{Good+}, \alpha_{Good-}, \alpha_{Bad+}, \alpha_{Bad-}, P, \beta$	6	RL
9	/	$\alpha, \gamma, \beta$	3	Bayesian
10	/	$\alpha, \gamma, w, \beta$	4	Bayesian

Table S1. Overview of tested models.

*Parameter estimation.* Parameter estimation was conducted using the maximum-likelihood approach, which finds the set of parameters that maximize the probability of the participant's trial-by-trial test phase choices given the model. Optimization was done by to minimizing the negative log-likelihood,  $-L$ , computed by:

$$-L = -\sum_{t=1}^T \ln (P_{choice}(t)) \quad [8]$$

where  $T$  denotes the total number of trials. Parameters were independently fitted to the test phase data for each participant using the Nelder-Mead optimization method. To avoid local minima in parameter fitting, optimization was initiated with 30 randomly selected start values. Model implementations and parameter fitting was done in *R* 3.5.1.

*Model comparison.* Model comparison was primarily based on the Akaike Information Criterion (AIC), a measure of goodness of fit of a model that penalizes complexity (Daw, 2011):

$$AIC = -2 \ln(L) + 2k \quad [9]$$

where  $-\ln(L)$  is the negative log-likelihood and  $k$  is the number of model parameters. A smaller AIC hence indicates a better model fit.

Model comparison was based on the sum AIC across participants. For simplicity, we present the  $\Delta AIC$ , which is the difference between model  $i$  and the best fitting model. We also

used AIC as an approximation to model evidence to compute the exceedance probability for each model, using Bayesian random effects model comparison. The exceedance probability expresses the probability that a given model is the most common model among the candidate models in the population (Stephan, Penny, Daunizeau, Moran, & Friston, 2009).

Table S2 shows the  $\Delta AIC$  for each experiment separately, as well as the combined  $\Delta AIC$ . Model 5, combining biased prior expectations and biased learning rates for each group, fit the data best in experiments 1-2 and 5. Experiment 3 was best fit by model 8, which included separate learning rates for positive and negative prediction errors (equation 1) from group A and B. However, the difference between model 5 and 8 in Experiment 3 was relatively small. To formally test for the reliability of the apparent difference between experiments, we used a random-effects approach. Specifically, we used a linear mixed model with AIC as the dependent variable, and participant as random factor to test the interaction between experiment and model. This approach showed a main effect of model,  $F(8, 3420) = 2.79$ ,  $p = .004$ , indicating that model 4 had significantly lower AIC than the other models. However, there interaction between model and experiment was not reliable,  $F(24, 3420) = 0.83$ ,  $p = .69$ , indicating that model 5 provided the best fit across experiments. In addition, Bayesian model comparison indicated that the posterior probability that all four experiments had the same model frequency was  $P = 0.96$ . Combining all experiments, we also find that the exceedance probability that model 5 was the most common among the candidate models was 1. Together, these results indicate that a combination of biased priors and biased learning rates best accounted for the influence of group descriptive information on instrumental learning across experiments.

Finally, we tested whether model fit differed between participants who identified as Republican or Democrat in Experiment 5. We found that the same model (#5) fit both Republican and Democrat participants (Table S3).

Model	1	2	3	4	5	7	7	8	9	10
<b>Experiment 1</b>	511	1140	213	175	0	211	252	228	670	496
<b>Experiment 2</b>	709	703	474	71	0	496	843	1141	486	505
<b>Experiment 3</b>	884	1536	196	199	70	163	114	0	205	199
<b>Experiment 5</b>	1344	2667	392	390	0	658	651	670	715	1812
<b>Combined</b>	3378	6046	1206	764	0	1458	1789	1968	2034	2942

*Table S2.*  $\Delta$ AIC by model. The table shows the difference in AIC, summed across participants, for each model relative to the best fitting model (with  $\Delta$ AIC = 0).

Model	1	2	3	4	5	6	7	8	9	10
<b>Democrats</b>	946	2122	276	292	0	538	491	490	447	1210
<b>Republicans</b>	398	546	116	98	0	120	160	180	268	602

*Table S3.*  $\Delta$ AIC by model and political party in Experiment 5. The table shows the difference in AIC, summed across participants, for each model relative to the best fitting model (with  $\Delta$ AIC = 0) for Democrat and Republican participants, respectively.

*Relation between model parameters and group-based bias.* To understand in more detail how the parameters of Model 5 related to the stereotype bias observed in choice behavior, we regressed the estimated model parameters (excluding the Softmax temperature  $\beta$ ) onto the degree of choice preference for positively-stereotyped group members in the test



phase (proportion of *Group A* choices). All model parameters were rank transformed and standardized to improve linearity and interpretability. We first conducted this analysis for Study 1-3. We found that both the prior  $P$  ( $\beta = 0.084$ ,  $SE = 0.012$ ,  $t = 6.62$ ,  $p < .0001$ ) and the learning rate parameter for the negatively-stereotyped group (“Group B”;  $\alpha_{\text{bad}}$ :  $\beta = -0.039$ ,  $SE = 0.013$ ,  $t = -3.7$ ,  $p = .003$ ) statistically predicted the degree of group preferences. In other words, a larger initial value difference between the groups (the prior), together with a lower learning rate for Group B, was associated with a stronger bias. The learning rate for the positive-stereotyped group (“Group A”) was not reliably related to bias ( $\alpha_{\text{good}}$ :  $\beta = -0.003$ ,  $SE = 0.011$ ,  $t = -0.33$ ,  $p = .74$ ).

We next conducted the same analysis separately for Study 5, because a different manipulation of the stereotype was used than in Studies 1-3: in Study 5, one group was described negatively, but no stereotype was given for the other group (see main text). This analysis again indicated that both the estimated prior ( $\beta = 0.097$ ,  $SE = 0.012$ ,  $t = 8.144$ ,  $p < .0001$ ) and the learning rate for negatively-stereotyped group (“Group B”;  $\alpha_{\text{bad}}$ :  $\beta = -0.032$ ,  $SE = 0.012$ ,  $t = -2.71$ ,  $p = .007$ ) uniquely predicted group-based choice preference. In addition, the learning rate for the positively-stereotyped group (“Group A”) was positively related to expressed bias ( $\alpha_{\text{good}}$ :  $\beta = 0.049$ ,  $SE = 0.012$ ,  $t = 4.18$ ,  $p < .0001$ )—a novel finding for this study that may reflect the lack of stereotype information for this group and stronger motivation to update from its members’ feedback. That is, a larger prior, a higher learning rate for the positively-stereotyped group and a lower learning rate for negatively-stereotyped group were associated with a stronger group-based choice preferences.

*Relation between model parameters and group-based bias.* Finally, to understand in more detail how the parameters of model 5 were related to the behavioral bias, we regressed the estimated model parameters (excluding the Softmax temperature  $\theta$ ) on the bias in the test phase (proportion *Group A* choices). All model parameters were rank transformed to improve linearity. We found that both the prior  $P$  ( $\beta = 0.087$ ,  $SE = 0.009$ ,  $t = 9.97$ ,  $p < .0001$ ) and the learning rate parameters ( $\alpha_{\text{good}}$ :  $\beta = 0.023$ ,  $SE = 0.008$ ,  $t = 2.85$ ,  $p = .005$ ,  $\alpha_{\text{bad}}$ :  $\beta = -0.032$ ,  $SE = 0.009$ ,  $t = -3.7$ ,  $p = .0002$ ) were related to the degree of bias. In other words, a larger initial value difference between the groups, together with a higher learning rate for Group A and a lower learning rate for Group B was associated with a stronger bias. We pooled the data from all four direct learning experiments (Study 1 – 3 & 5) for these analyses.

### Supplementary References

- Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of social issues*, 55, 429-444.
- Daw, N. D. (2011). Trial-by-trial data analysis using computational models. *Decision making, affect, and learning: Attention and performance XXIII*, 23.
- Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton trilogy revisited. *Journal of personality and social psychology*, 11, 1139-1150.
- Doll, B. B., Jacobs, W. J., Sanfey, A. G., & Frank, M. J. (2009). Instructional control of reinforcement learning: a behavioral and neurocomputational investigation. *Brain research*, 1299, 74-94.
- Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, 273, 1399-1402.
- Park, B., & Rothbart, M. (1982). Perception of out-group homogeneity and levels of social categorization: Memory for the subordinate attributes of in-group and out-group members. *Journal of personality and social psychology*, 42, 1051.
- Pettigrew, T. F. (1979). The ultimate attribution error: Extending Allport's cognitive analysis of prejudice. *Personality and social psychology bulletin*, 5, 461-476.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, 2, 64-99.
- Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *Neuroimage*, 46, 1004-1017.