Internalization of societal stereotypes as individual prejudice

Benjamin S. Stillerman\textsuperscript{1}, Björn R. Lindström\textsuperscript{2}, David T. Schultner\textsuperscript{2}, Leor M. Hackel\textsuperscript{3}, Damaris R. Hagen\textsuperscript{2}, Nils B. Jostmann\textsuperscript{2}, & David M. Amodio\textsuperscript{*1, 2}

\textsuperscript{1}New York University, \textsuperscript{2}University of Amsterdam, \& \textsuperscript{3}University of Southern California

Correspondence to:
David M. Amodio
Department of Psychology
New York University
6 Washington Place
New York, NY 10003
David.amodio@nyu.edu

Keywords: Prejudice, Stereotyping, learning, social cognition, reinforcement
Abstract

How do stereotypic messages about social groups become internalized in an individual’s own preferences and behaviors? We proposed that exposure to such group messages influences how we learn from individual group members in subsequent direct interactions. In a series of experiments, participants interacted with players from two groups, described with positive or negative stereotypes, in a monetary sharing game. Although the actual sharing rates of individual players were equated between groups, stereotypic messages biased how participants learned from and interacted with individual players—a pattern not evident in deliberative decisions or explicit beliefs regarding players’ sharing rates. Computational models revealed that this group preference reflected a combination of biased priors and biased reward updating: stereotypes set participants’ initial expectancies about group members’ behaviors and impaired their ability to learn through direct interaction and feedback. Next, we found that the expression of this bias in a direct learner’s behavior is unwittingly propagated to naïve observational learners, who come to show bias in their own behaviors. In a final set of experiments, we demonstrated this pattern of internalization and propagation when a stereotypic message is attributed to a political leader (Donald Trump), even among participants with opposing political views. These findings offer the first mechanistic account of how explicit messages about group stereotypes and inequalities may be internalized in an individual’s mind as prejudice, from where it is expressed implicitly and propagated to others.

(232 words)
How do explicit stereotypical messages about social groups become internalized in an individual’s own preferences and behaviors? When a politician refers to a group as “criminals and rapists,” people may dismiss the epithets as mere rhetoric. Yet such messages, conveyed explicitly, are nevertheless encoded in the recipient’s memory. We asked whether this kind of information, once encoded in memory, can shape the way people subsequently perceive and interact with members of the targeted group, leading to the internalization of prejudice in one’s own response preferences and its unwitting propagation to people who observe these responses.

To understand how internalized prejudices could form in response to explicit messages, we considered the interplay of learning mechanisms involved in the processing of an explicit message and the formation of implicit choice preferences, respectively. Explicit messages, such as epithets about a social group, are processed and encoded in semantic memory, which represents general knowledge about the world that is reportable and, when considered inaccurate, discounted (Collins & Quillian, 1969; Gilbert, 1991). By comparison, internalized social choice preferences may be characterized in terms of reward associations (Hackel, Berg, Lindström, & Amodio, 2019; Kurdi, Gershman, & Banaji, 2019; Lindström, Selbing, Molapour, & Olsson, 2014), encoded incrementally across repeated interactions through the process of reinforcement learning (Daw, O’Doherty, Dayan, Seymour, & Dolan, 2006). This type of feedback-based instrumental learning is supported by dopaminergic activity in the striatum (O’Doherty et al., 2004; Schultz, Dayan, & Montague, 1997; Yin & Knowlton, 2006) and is expressed in an individual’s personal choice behaviors (Delgado, Frank, & Phelps, 2005; Shohamy & Wagner, 2008). That is, whereas one’s knowledge of a group stereotype may not be personally endorsed (Devine, 1989), instrumentally-learned tendencies toward group members would reflect one’s internalized choice preferences (Kurdi et al., 2019).
We hypothesized that the transmission of prejudice from an explicit message to an internalized preference involves the interplay of these two memory processes; that is, following exposure to an explicit semantic message about a social group, this information will influence instrumental learning in subsequent interactions with individual members of the group. This effect could occur by biasing initial expectations about a group member’s behavior and by altering how the reward value of a group member’s behavior is interpreted and updated. We propose that through this process, explicitly-conveyed information about social stereotypes and structures may become internalized within an individual as biased response tendencies toward group members. Moreover, once internalized and expressed in behavior, these biases may be inadvertently transmitted to others who observe this behavior, suggesting a model of how stereotypic social messages may be heard by one individual and then unwittingly spread to others.

To test this hypothesis, we conducted a series of experiments in which research participants learned they would interact with people from two different communities located in different regions via an online game. These two groups, labeled as Group A and B (counterbalanced) in the task ostensibly to maintain their anonymity, were described using positive and negative stereotypes commonly associated with White and Black Americans, with attributes indicating relative socioeconomic status, crime prevalence, and trait attributes (Devine & Elliot, 1995; see SI). 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 played 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 the present version of the task, 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%), with sharing rates equated between the 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 the training phase, 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).

Figure 1. Schematic of the learning task training phase. (A) Participants 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 and shirt color), and gender was counterbalanced across participants. (B) In the training phase, participants chose between players...
(group members) with reciprocal reward rate (e.g., 70% and 30%) and (C) and received immediate reward feedback, as shown in this sample trial. In the test phase, 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, in which they viewed and selected between all possible pairings of Group A and B members. This design provided a fine-grained behavioral assessment of learned reward associations with each member of the two groups (Frank et al., 2004). Feedback was not given after test phase choices; participants received payout for their choices at the task’s conclusion. Training included 160 trials; test included 96 trials.

In an initial laboratory experiment (Study 1; N = 61), we tested whether stereotypic group descriptions influenced participants’ choices of individual players, despite equivalent sharing rates between groups—the hallmark of group-based prejudice. Test phase behavior showed that participants learned the general pattern of rewards, choosing players with higher sharing rates on average ($B = 2.68$, $SE = 0.19$, Wald $z = 14.43$, $p < .001$). However, their choices were also significantly biased by players’ group membership ($B = 0.52$, $SE = 0.06$, Wald $z = 9.33$, $p < .001$; Figure 2A), despite their direct experience with individual’s actual sharing rates (results were equivalent during the training phase; see SI). Given participants’ explicit goal to learn each players’ sharing rate and to maximize earnings, this group effect represents an indirect (i.e., implicit) expression of bias and suggests, as hypothesized, that explicit information about the groups transferred to participants’ implicit personal preferences for individual group members.
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 predictions from the best fitting RL model, which combined biased group-based priors and learning rates (see text for details).

To more directly characterize the effect of this response bias as instrumental learning, and to probe the process through which this bias occurred, we fit behavior to computational models of reinforcement learning. We considered two potential mechanisms of bias in these models: First, positive or negative group descriptions may bias initial expectances, represented in our model as priors regarding the value of each group. Second, these messages could bias learning rates—the degree to which reward associations are updated in response to new information.
Comprehensive tests of these reinforcement learning models (as well as Bayesian learning models) indicated that a hybrid model, which includes both processes, best explained the observed behavior (see SI for model comparisons). According to this model, the group bias was the result of initial expectancies combined with impeded updating of value when interacting with members of the negatively-described group. This finding supports the idea that explicit group messages can indeed become internalized in an individual’s value representations of a group through the process of interactive reinforcement learning.

This pattern of group message internalization was replicated in two additional studies with online samples (Study 2: \(N = 62\); Study 3: \(N = 87\)). In both, explicit group descriptions significantly influenced participants’ choice preferences in both the training and test phases (Study 2 test phase: \(B = 0.79, SE = 0.06, \text{Wald } z = 13.86, p < .001\); Study 3 test phase: \(B = 0.48, SE = 0.05, \text{Wald } z = 9.58, p < .001\); see SI), despite equivalent average sharing rates between groups and participants’ explicit goal to learn about individual players in order to maximize earnings. Computational models again indicated that this effect best reflected a combination of biased priors and learning rates, such that the group messages set initial expectancies that were never sufficiently updated in the learning process (see timecourse data in SI).

In Study 3, we additionally asked whether the internalization of explicit group biases in participants’ choice preferences would be expressed in trust decisions. Following the learning task, participants completed separate single-round trust games with each player from the learning task. In these games, participants could entrust a portion of their points to a player; this amount would be tripled, and the player could then return a portion of this larger amount to the participant (converted to cash at study conclusion, see Bohnet & Zeckhauser, 2004). We expected participants to entrust points based on their prior learning experience with each player.
This task allowed us to examine whether the group bias evident in choice behaviors would also emerge in a more deliberative form of social decision making. On average, participants entrusted more money to players with higher sharing rates, $B = 4.87, \ SE = 1.90, t(692) = 2.57, p = .010, 95\% \ CI [1.15, 8.59]$ (Figure 3a). However, trust decisions were not moderated by group membership ($B = -1.63, \ SE = 1.41, t(692) = -1.16, p = .249, 95\% \ CI [-4.38, 1.14]$), nor the Reward rate x Group interaction ($B = 2.00, \ SE = 2.69, t(692) = 0.75, p = .457, 95\% \ CI [-3.26, 7.26]$), suggesting that the social bias evident in participants’ choice behavior did not emerge in their deliberative decisions.

![Figure 3.](image)

**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.

An additional question of Study 3 was whether participants subjectively perceived a group difference in reward rates which was apparent in their choice behaviors. Prior studies have shown that probabilistic feedback-based learning tasks, such the task used here, can produce learning expressed in behavioral choices despite poor declarative knowledge of reward rates (Knowlton, Mangels, & Squire, 1996). To assess the extent that the group bias observed in
behavior operated independently of participant’s explicit representation of player reward values, we assessed participants’ belief about each player’s sharing rate following the trust game. For each player in the learning task, participants indicated the number of times they believed the player would share in 100 hypothetical interactions. Although participants’ explicit beliefs of reward rates showed a degree of group bias ($B = 3.88$, $SE = 1.90$, $t(683) = 2.04$, $p = .042$), their reward rate estimates were poorly calibrated to actual rates (Fig 3b). A regression analysis in which these explicit reward estimates were covaried, and thus statistically removed, continued to demonstrate a significant effect of group bias on participants’ choice behavior, $B = 0.21$, $SE = 0.05$, Wald $z = 3.97$, $p < .001$. These results suggest that the group bias exhibited in participants’ choice behavior did not merely reflect an explicit belief about group differences, but an internalized form of instrumental reward association.

Together, Studies 1–3 demonstrate a novel mechanism through which explicit information about social stereotypes may become internalized in an individual’s personal group-based preferences. Once encoded in instrumental value representations of group members, these preferences were expressed indirectly in subsequent social decisions, irrespective of explicit beliefs regarding sharing rates.

Next, to begin to understand how explicit social messages may be implicitly propagated across a population, we asked whether such messages, once internalized and expressed in behavior, could be inadvertently transmitted to others who observe this behavior. Hence, in a preregistered fourth study, we tested whether the mere observation of another person’s learned social preferences would also become internalized in the observer. Participants in Study 4 ($N = 124$) were invited to play the money sharing game used in Studies 1–3. However, instead of learning directly from player feedback in a learning phase, participants observed the training-
phase choice behavior and feedback of a prior participant. Observers’ explicit goal was to learn players’ sharing rates in order to improve their own chances of winning money in a subsequent test phase with the same players (Olsson, Knapska, & Lindström, 2020). Each Study 4 participant (observer) viewed the learning phase behavior and reward feedback of a different participant from Study 2 (herefore referred to as a demonstrator) in a yoked design. Observers were told only that the players in the game represented two different social groups; they did not receive the group stereotype descriptions provided to demonstrators.

Did the mere observation of demonstrators’ behavior and feedback induce a group preference in the observers? Indeed, observers exhibited a significant group bias in their own test phase choices, in line with the original explicit message conveyed to demonstrators ($B = 0.32$, $SE = 0.04$, Wald $z = 8.03$, $p < .001$): members of the group described positively to demonstrators were preferred by observers across levels of reward—an pattern of bias that was effectively transmitted despite the subtlety of demonstrators’ choice preferences and the lack of group descriptions given to observers, and despite equivalent reward feedback from players. As in Study 3, this group bias remained even after adjusting for Study 4 participants’ explicit beliefs regarding payer reward rates ($B = 0.26$, $SE = 0.04$, Wald $z = 6.52$, $p < .001$). These findings suggest that an explicit social message affects not only direct learners in social interactions, but also subsequent observers of these interactions who had not been exposed to group messages, suggesting a potential mechanism for implicit bias propagation.
Figure 4. Observer choice behavior for the test phase of Study 4. Observers 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.

Finally, having demonstrated that explicit group descriptions can be internalized in an individual’s own choice preferences through the process of instrumental learning, and then propagated to others through mere observation, we returned to the question that began this report: how might a politician’s negative rhetoric about a social group become internalized, even among people who object to the politician and dismiss his or her 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, participants (N = 200) 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 were told that President Trump had not commented on the second group, from which the other players came, and thus the second group in this study constituted a neutral control group. Following the training and test phases, as in Study 1, participants reported their political party identification (Democrat, $N = 162$ or Republican, $N = 38$) and their general agreement with Donald Trump’s positions on a scale (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 social group descriptions of Donald Trump influence participants choice preferences? Indeed, participants were more likely to avoid members of the group disparaged by President Trump, despite the equivalent average sharing rates between groups ($B = 0.40$, $SE = 0.03$, Wald $z = 12.78, p < .001$). Moreover, 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 $z = 4.58, p < .001$). Notably, while our sample included fewer self-identified Republicans than Democrats, our primary interest was in the behavior of Democrats, who demonstrated an internalization of the group bias despite their disagreement with Trump’s messages.
Figure 5. Solid lines depict test phase choice behavior for Study 5 (Panels A and B depict Democrat and Republican choices, respectively). Participants of both political parties demonstrated successful learning of reward contingencies and a group bias, although the effect was enhanced among Republicans. 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 best fitting RL model, which combined biased group-based priors and learning rates (see text for details).

Computational models again revealed that this pattern reflected a combination of biased priors and biased learning, reflecting lower initial expectancies combined with altered updating for members of the group disparaged by Trump—a pattern that reflects an internalized devaluation of group members (see SI). This model provided the best fit of the data even when applied separately to Republican and Democrat participants. Together, these results suggest that negative group messages, conveyed explicitly by a politician, can be internalized through the process of direct instrumental interactions with individual group members, and that this internalization process may occur despite strong opposition to the politician.

As a last step, we tested whether these group biases, conveyed by a political leader and internalized by direct learners regardless of their political party, could spread to others through mere observation of a direct learner’s behavior, as in Study 4. Participants in Study 6 observed
past 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 who were described neutrally to
Study 5 participants over those 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, participants who observed Republican demonstrators showed a larger group bias ($B
= 0.35, SE = 0.05$, Wald $z = 6.36, p < .001$) than those who observed Democrats ($B = 0.21, SE =
0.04$, Wald $z = 5.28, p < .001$). Nevertheless, these results demonstrate that the bias was
conveyed by the behavior of both Democrat and Republican demonstrators, despite its origin in
the message of a Republican president.

In this research, we asked whether explicitly-conveyed stereotypes about a group can
influence how we learn about individual group members in a way that produces an internalized
social bias. Across six studies, we found that, indeed, positive and negative group descriptions
regarding stereotypes and structural inequalities, conveyed explicitly, biased the process of
incremental reward reinforcement learning in direct interactions with group members. This
process involved the interplay of two modes of learning—semantic learning of explicit
conceptual descriptions and the instrumental learning of reward values in direct interactions—
and represents the transmission of an explicit group description to an individual’s internalized
choice preference and biased behavior. This transmission was evident despite participants’
limited declarative knowledge of reward rates, and it was expressed despite their explicit goal to
earn money and, in Study 5, their disagreement with the messenger—a pattern suggesting the indirect expression of group preference that characterizes implicit bias. Moreover, the resulting pattern of bias in participants’ choice behavior was transmitted to unwitting observers who, upon viewing this behavior with no knowledge of group differences or an intention to discriminate, expressed this bias in their own decision making. Together, these studies reveal a process through which explicit messages about a social group maybe internalized as implicit bias and then transmitted to unsuspecting others.

Our computational models further illuminated the mechanisms through which explicit messages about social groups can bias direct learning and choice. These models indicated two patterns of influence: a bias in the priors which anchored expectations of a group member’s behavior and, independently, a bias in the extent to which reward representations of a group is updated in the face of new information. Whereas prior computational work examined the effect of specific selection rules on reinforcement learning (Doll, Jacobs, Sanfey, & Frank, 2009), our findings reveal the effect of a broad conceptual description on reward-based choice behaviors—a novel demonstration of interactive memory systems in the context of social cognition (Amodio, 2019). This computational analysis provides a new theoretical account of how implicit prejudices form from environmental influences and, by identifying specific sources of bias, may suggest unexplored strategies for reducing the impact of divisive rhetoric and social stereotypes on an individual’s intergroup attitudes and behaviors.

More broadly, our findings suggest that explicit messages favoring or derogating social groups are more than mere words; exposure to a biased political messages can shape an individual’s subsequent experiences with members of the group, perhaps without their knowledge, in a way that confirms the message and spreads it to others. This process—whereby
social stereotypes are transformed into internalized prejudices—may similarly explain how systemic biases, such as institutional inequality, may be transmitted from social structures to the minds of individuals (Payne, Vuletich, & Lundberg, 2017). As society continues to grasp and gauge the impact of 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 implicit bias in the individual, these results promise to identify new approaches to assessing and reducing the impact of stereotypes and system racism on individuals’ prejudices and intergroup relations.
References


Supplemental Information

Table of Contents

Study 1 ................................................................................................................................. 21
   Method .............................................................................................................................. 21
   Results .............................................................................................................................. 24
Study 2 ................................................................................................................................. 25
   Method .............................................................................................................................. 25
   Results .............................................................................................................................. 26
Study 3 ................................................................................................................................. 27
   Method .............................................................................................................................. 27
   Results .............................................................................................................................. 28
Study 4 ................................................................................................................................ 30
   Method .............................................................................................................................. 30
   Results .............................................................................................................................. 32
Study 5 ................................................................................................................................ 34
   Method .............................................................................................................................. 34
   Results .............................................................................................................................. 35
Study 6 ................................................................................................................................ 37
   Method .............................................................................................................................. 38
   Results .............................................................................................................................. 39
Computational modeling .................................................................................................... 42
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; $M_{age} = 21.56$ years, $SD_{age} = 5.20$ years).

Procedure. Upon arrival to the lab, 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 stereotypes of White and Black Americans, respectively (Devine & Elliot, 1995), and which correspond to common stereotypes to White (native) and Moroccan Dutch immigrants.

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.

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. Female avatars and schematic of reward probabilities.](image)

Next, participants completed the main choice 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). 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, without feedback. However, 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 that 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, \( B = 2.68, SE = 0.19, \text{Wald } z = 14.43, p < .001 \). 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 x 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.

**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).

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 an explicit belief, in which they typed their estimate of the reward rate for each player (“How many times out of a hundred would this player share with you?”). This form of response was designed to most directly assess declarative semantic knowledge without picking up on motor or affective tendencies that might be associated with striatal-based value representations. 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].

Explicit beliefs. Participants’ explicit reward estimates were submitted to a linear regression, with actual player reward rate and player group as predictors. Explicit belief 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. Explicit belief was 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 explicit beliefs in addition to the relatively strong effect effects observed on choice behavior.

To test whether the effects of group membership on choice behavior could be explained by explicit beliefs, or whether the group bias instrumental learning operated independently of explicit beliefs, we conducted an analysis in which player reward rates, group membership, and explicit belief estimates were each included in the main multilevel model predicting test phase choice. We found that while explicit beliefs 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 explicit (declarative) 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 preregistered 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; \(M_{\text{age}} = 19.5\) years, \(SD_{\text{age}} = 1.35\) years).

**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. Afterwards, 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?”). Next, 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 differs slightly from our preregistered plan, which was to yoke Study 4 participants to all Study 2 participants collected ($N = 78$), rather than only
the participants included in the final Study 2 analysis. The pattern of results was the same using this sample \((N = 156)\), with effect size estimates almost identical to those reported with the smaller sample (test phase learning effect: \(B = 1.49, SE = 0.11, \text{Wald} \ z = 13.36, p < .001\); group effect: \(B = 0.39, SE = 0.03, \text{Wald} \ z = 11.25, p < .001\); interaction: \(B = -0.17, SE = 0.18 \text{ Wald} \ z = -1.13, p = .258\)).

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 \((B = 0.28, SE = 0.09, \text{Wald} \ z = 3.21, p = .001\). This demonstrates that participants exposed to more biased demonstrators themselves tended to be more biased.

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, \(B = 0.01, SE = 0.001, \text{Wald} \ z = 20.22, p < .001\), significant effects remained for actual reward rates, \(B = 1.69, SE = 0.13, \text{Wald} \ z = 13.09, p < .001\), and group membership, \(B = 0.26, SE = 0.04, \text{Wald} \ z = 6.52, p < .001\).
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).

**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, \( B = 1.98, SE = 0.10, \text{Wald } z = 19.40, p < .001 \), 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, $B = -0.03, SE = 0.14$ 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: $B = 0.31, SE = 0.04, \text{Wald } z = 8.14, p < .001$; Republicans: $B = 0.62, SE = 0.06, \text{Wald } z = 10.83, p < .001$). 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 estimates 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), $B = 0.03, SE = 0.01, \text{Wald } z = 42.07, p < .001$, actual player reward rates, $B = 0.90, SE = 0.11, \text{Wald } z = 8.10, p < .001$, and group membership ($B = 0.39, SE = 0.03, \text{Wald } z = 11.47, p < .001$) 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 \( N = 189 \).

*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 $z = 2.34$, $p = .022$).
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 ($B = 0.02, SE = 0.001, \text{Wald } z = 31.26, p < .001$), the influence of actual reward rates ($B = 1.97, SE = 0.11, \text{Wald } z = 18.20, p < .001$) and group membership ($B = 0.19, SE = 0.03, \text{Wald } z = 5.70, p < .001$) 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 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:

$$Q_{i}^{t+1} = Q_{i}^{t} + \alpha(R^{t} - Q_{i}^{t})$$

[1]
where \( Q \) 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

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

[2]

where \( \beta \) (0.01 < \( \beta \) ≤ 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 ~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 \( Q_{Good}^{t=0} = prior, Q_{Bad}^{t=0} = -prior \). 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 \( P > 0 \). It should be noted that 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 2).

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., $Ca_{positive}\delta_+, (\frac{\alpha_{negative}}{C})\delta_-$), and vice versa for the bad group (i.e., $\frac{\alpha_{positive}}{C}\delta_+, Ca_{negative}\delta_-$) (see Table S1 for overview).

Figure S3 depicts participant choice over time in all modeled studies (Study 1 – 3 and 5) to descriptively illustrate the presence of an initial prior bias as well as a persisting tendency to choose Group A.
Figure S3. Participant likelihood of choosing Group A (over Group B) in the training phase by trial number, for Studies 1–3 and 5 combined (top panel) and individually. Likelihoods are smoothed with a 10-trial moving average.

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:
\[ \alpha_{t+1} = \alpha_t + pos \]  \quad [3] \\
\[ \beta_{t+1} = \beta_t + neg \]  \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:

\[ \text{mode}_i = \frac{a_i - 1}{a_i + \beta_i - 2} \]  \quad [5]

We evaluated two versions of this model. In the first version (model 8), 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 \( a \). More evidence is required to counteract a precise prior. If \( a_{\text{Good}} \) is higher than \( a_{\text{Bad}} \), the model is biased to select group A. The second model (model 9) 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

\[ \alpha_{t+1} = \alpha_t + w pos \]  \quad [6] \\
\[ \beta_{t+1} = \beta_t + \frac{1}{w} neg \]  \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).

<table>
<thead>
<tr>
<th>Model #</th>
<th>parameters</th>
<th># parameters</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\alpha, \beta$</td>
<td>2</td>
<td>RL</td>
</tr>
<tr>
<td>2</td>
<td>$\alpha, \beta, P$</td>
<td>3</td>
<td>RL</td>
</tr>
<tr>
<td>3</td>
<td>$\alpha_{Good}, \alpha_{Bad}, \beta$</td>
<td>3</td>
<td>RL</td>
</tr>
<tr>
<td>4</td>
<td>$\alpha_{Good}, \alpha_{Bad}, P, \beta$</td>
<td>4</td>
<td>RL</td>
</tr>
<tr>
<td>5</td>
<td>$\alpha_{Good}, \alpha_{Bad}, C, \beta$</td>
<td>4</td>
<td>RL</td>
</tr>
<tr>
<td>6</td>
<td>$\alpha_{Good}^+, \alpha_{Good}^-, \alpha_{Bad}^+, \alpha_{Bad}^-$</td>
<td>5</td>
<td>RL</td>
</tr>
<tr>
<td>7</td>
<td>$\alpha_{Good}^+, \alpha_{Good}^-, \alpha_{Bad}^+, \alpha_{Bad}^-, P, \beta$</td>
<td>6</td>
<td>RL</td>
</tr>
<tr>
<td>8</td>
<td>$\alpha, \gamma, \beta$</td>
<td>3</td>
<td>Bayesian</td>
</tr>
<tr>
<td>9</td>
<td>$\alpha, \gamma, w, \beta$</td>
<td>4</td>
<td>Bayesian</td>
</tr>
</tbody>
</table>

*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:

$$-L = - \sum_{t=1}^{T} \ln (P_{\text{choice}}(t))$$

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):

$$AIC = -2 \ln(L) + 2k$$

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 ΔAIC for each experiment separately, as well as the combined ΔAIC. Model 4, 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 7, which included separate learning rates for positive and negative prediction errors (equation 1) from group A and B. However, the difference between model 4 and 7 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 4 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 4 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 (#4) fit both Republican and Democrat participants (Table S3).

<table>
<thead>
<tr>
<th>Model</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment 1</strong></td>
<td>511</td>
<td>213</td>
<td>175</td>
<td>0</td>
<td>211</td>
<td>252</td>
<td>228</td>
<td>670</td>
<td>496</td>
</tr>
</tbody>
</table>
Table S2. ΔAIC by model. The table shows the difference in AIC, summed across participants, for each model relative to the best fitting model (with ΔAIC = 0).

<table>
<thead>
<tr>
<th>Model</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democrats</td>
<td>946</td>
<td>276</td>
<td>292</td>
<td>0</td>
<td>538</td>
<td>491</td>
<td>490</td>
<td>447</td>
<td>1210</td>
</tr>
<tr>
<td>Republicans</td>
<td>398</td>
<td>116</td>
<td>98</td>
<td>0</td>
<td>120</td>
<td>160</td>
<td>180</td>
<td>268</td>
<td>602</td>
</tr>
</tbody>
</table>

Table S3. ΔAIC 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 ΔAIC = 0) for Democrat and Republican participants, respectively.

Relation between model parameters and group-based bias. Finally, to understand in more detail how the parameters of model 4 were related to the behavioral bias, we regressed the estimated model parameters (excluding the Softmax temperature $\beta$) 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 ($a_{\text{good}}: \beta = 0.023, SE = 0.008, t = 2.85, p = .005$, $a_{\text{bad}}: \beta = -0.032, SE = 0.009, t = -3.7$, $p = .0002$) 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


