PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES

A role for self-regulatory control in the absence of ingroup preference

Samuel A. W. Klein[a,1], Tessa E. S. Charlesworth[b], Daniel W. Heck[c], Mahzarin R. Banaji[d], and Jeffrey W. Sherman[e]

Affiliations are included on p. 8.

Edited by B. Keith Payne, University of North Carolina at Chapel Hill, Chapel Hill, NC; received January 5, 2026; accepted August 17, 2026 by Editorial Board Member Nora S. Newcombe

One of psychology's most ubiquitous findings is that of robust ingroup preference, an explicitly stated human preference for the groups to which we belong. The advent of implicit measures of social cognition introduced a paradoxical result: Compared to members of higher-status groups, members of lower-status groups often show no implicit ingroup preference, sometimes even showing outgroup preference, a result rarely observed on explicit measures. Prevailing theories have provided macro-level accounts, often in terms of system-justifying tendencies, but the various cognitive processes that could produce such an absence of ingroup preference among lower-status groups have not been directly decomposed. In this paper, we address this gap via formal process modeling applied to data from 5.7 million respondents collected across 16 years and 6 social categories. We find that, among lower-status groups, ingroup-affirming evaluations are accompanied by more pronounced self-regulatory control than outgroup-affirming evaluations, a regulatory asymmetry absent among higher-status groups. For example, evaluations affirming Black and overweight ingroups were accompanied by heightened self-regulatory control, whereas a comparable self-regulatory pattern was not observed for evaluations affirming White and nonoverweight ingroups. The evidence points to a previously unrecognized role for self-regulatory control in the absence of ingroup preference, suggesting that the cognitive imprint of social hierarchies may extend beyond evaluative associations to the regulatory processes that govern their expression.

implicit social cognition | Implicit Association Test | intergroup bias | outgroup bias | computational cognitive modeling

Significance

While humans generally exhibit a robust preference for their own social groups, members of lower-status groups frequently lack this preference on implicit measures, a pattern with consequences for how social hierarchies are maintained. To investigate the cognitive mechanisms underpinning this phenomenon, we employed formal process modeling of 5.7 million implicit bias tests across six social categories (age, disability, race, skin tone, sexuality, and weight). We find that, among lower-status groups, ingroup-affirming evaluations are accompanied by more pronounced self-regulatory control than outgroup-affirming evaluations, a regulatory asymmetry absent among higher-status groups. These findings suggest that the cognitive imprint of social hierarchies may extend beyond the evaluative associations people acquire to the regulatory processes that govern their expression.

Ingroup preference is arguably the most widely documented finding in the science of intergroup behavior (1). It manifests in a vast array of biased behavior, including cooperation and resource allocation (2, 3), trust and lending decisions (4), helping in emergencies (5), perception and memory (6, 7), empathic responses to pain (8, 9), and the application of moral judgments and punishment (10, 11). Ingroup preference emerges even within 'minimal group' paradigms, in which group membership is arbitrary (12), aligning with the perspective that ingroup love is universal and highly adaptive for survival and social coordination (13, 14).

Yet a recurrent and puzzling counter-pattern has also emerged over the years. In the 1940s, Kenneth and Mamie Clark (15) conducted pioneering studies in which they offered young Black children a choice, a White doll or a Black doll. Black children consistently chose White dolls over Black dolls, the first empirical result demonstrating an absence of ingroup preference. By the 1990s, numerous cases emerged of ingroup bias missing from the behaviors of lower-status group members (16). It eventually became clear that a lack of preference for lower-status ingroups, or even outright preference for higher-status outgroups, was most robustly detected through implicit methods, designed to capture evaluations (attitudes) and stereotyping (group-based beliefs) that operate relatively outside conscious awareness, control and intention.

Members of higher-status groups (e.g., White, straight, young, thin) consistently exhibit strong implicit ingroup preference (i.e., ingroup-affirming behavior; 17–19). Members of lower-status groups, however, frequently do not. For example, Black and Hispanic Americans have each demonstrated implicit outgroup bias (20–22), and children from various lower-status groups show outgroup bias to the extent that they perceive status differences between groups (23). In contrast, explicit measures of evaluation and stereotyping more consistently reveal ingroup preference, regardless of the respondent's status within society (18).

Internalization of Culturally Dominant Attitudes. Prevailing theory explains this contradiction of ingroup preference through mechanisms of cultural learning. Like their higher-status counterparts, lower-status group members presumably are inherently motivated to affirm their ingroup (24, 25). However, unlike their higher-status

The authors declare no competing interest.

This article is a PNAS Direct Submission. B.K.P. is a guest editor invited by the Editorial Board.

Copyright © 2026 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]To whom correspondence may be addressed. Email: samuel.klein@columbia.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2600311123/-/DCSupplemental.

Published September 9, 2026.

Downloaded from https://www.pnas.org by COLUMBIA UNIVERSITY on September 9, 2026 from IP address 128.59.177.134.

counterparts, lower-status groups must simultaneously contend with widely shared negative beliefs about their group in society. One prominent account of how this tension is resolved draws on system justification theory (SJT; 16, 18, 26), which posits that a competing motivation to view broader social systems as just and legitimate leads lower-status groups to internalize and justify culturally dominant attitudes that disadvantage them—the internalization of both negative ingroup stereotypes and positive outgroup stereotypes.[*]

The Cognitive Processes of Implicit (Outgroup) Preference. Although compelling as a macro-level account, SJT's proposed internalization of culturally dominant attitudes does not specify the cognitive mechanisms that produce an absence of ingroup preference. Associative learning offers one plausible cognitive pathway: lower-status groups inhabit a stratified social ecology where media and everyday social interaction frequently link their group with negativity and higher-status groups with positivity (27, 28). Through statistical learning, such exposure could establish outgroup-affirming associations (29, 30). Such learning may produce various outgroup-affirming associations, ranging from positive outgroup evaluations (31) to distinct threat-related responses (32, 33). Whether cognitive processes beyond such associative acquisition contribute to the absence of ingroup preference remains unclear.

Formal process modeling has revealed that responses on implicit methods reflect a constellation of processes, not merely the activation of evaluative associations (34–37). Among those additional cognitive components is self-regulatory control, or the capacity to prevent evaluative associations from driving biased responses (38, 39). Self-regulatory control is empirically dissociable from both the activation of evaluative associations and general task accuracy, and it is readily engaged even during tasks designed to limit controlled processing (38–41). Moreover, inhibitory responses can be conditioned to activate much like evaluations, through repeated pairing of stimuli with cues to inhibit otherwise typical responses (42–45).

Theoretically, whereas the absence of ingroup preference can be attributed to the evaluative associations people acquire, it could also reflect the regulatory control directed at those evaluations. Self-regulatory control is well established as a component of implicit intergroup behavior, yet whether it contributes to the absence of ingroup preference among lower-status groups remains unexplored. We therefore ask not only whether outgroup-affirming evaluations are more prevalent among lower-status groups, but whether ingroup-affirming evaluations are accompanied by greater regulatory control than outgroup-affirming evaluations.

Current Research. To decompose the processes underwriting status differences in ingroup preference, we applied the Quadruple Process model (QPM; 38) to responses on 5.7 million Implicit Association Tests (IATs; 46) from the Project Implicit demonstration website.[†] Data were collected over 16 years and across 6 social categories: age (elderly/young), disability (disabled/able), race (Black/White), skin tone (dark/light), sexuality (gay/straight), and weight (fat/thin). First, we analyzed whether outgroup-affirming (vs. ingroup-affirming) implicit bias is more prevalent among lower-status groups (elderly, disabled, Black, dark-skinned, gay, fat) than higher-status groups (young, able, White, light-skinned, straight, thin). We then isolated the contributions of the two key processes and subsequently tested 1) whether outgroup-affirming associations are more readily activated among lower-status groups and 2) the question of whether ingroup-affirming evaluations are accompanied by greater regulatory control than outgroup-affirming evaluations among lower-status groups.

Results

Lower-Status Groups' Disproportionate Levels of Outgroup-Affirming Behavior. Before turning to the QPM analyses, we first present the overall behavioral pattern in the IAT, a dual-categorization task designed to capture the ease with which respondents can sort social groups (e.g., elderly, young) and valenced attributes (e.g., good, bad). We relied on its most common metric of biased responding: the D-score. The IAT D-score quantifies the relative difference in response latencies between IAT block pairings (47). Fig. 1 displays the distribution of IAT D-scores, separately for higher-status and lower-status group members. We computed the D-scores such that positive values indicate faster responding on trials when the respondent's ingroup is paired with positive evaluative words (i.e., ingroup preference); negative D-scores indicate the reverse (i.e., outgroup preference). Descriptive and inferential D-score analyses across group status and attitude domains are reported in the *SI Appendix*.

Higher-status groups exhibited clear implicit ingroup preference across all six domains, with the vast majority of respondents showing positive D-scores. Lower-status groups, by contrast, showed substantially attenuated ingroup preference and even outgroup-affirming D-scores across domains. The strongest outgroup-affirming behavior was demonstrated by lower-status groups in the age, disability, and weight domains. Said another way, outgroup-affirming behavior was exhibited by 10 to 22% of higher-status respondents but 32 to 87% of lower-status respondents (Fig. 1). The QPM analyses that follow ask what constellation of cognitive processes underlies the attenuated or absent ingroup preference characteristic of lower-status groups.

Implementation of the QPM. We employed the QPM to isolate the cognitive mechanisms underpinning performance on the IAT. The QPM is a multinomial processing-tree model that decomposes the observed correct/incorrect responses on the IAT into four distinct latent processing parameters. The model assumes that a response on any given trial is the probabilistic outcome of a chain of cognitive states. To illustrate, consider a trial on an Age IAT where a respondent must sort an elderly target face using the same key assigned to sort "good" words. The QPM estimates the probability of four processes potentially occurring to produce responses on this trial:

Association activation. This parameter estimates the probability that the image of the elderly target triggers the activation of an evaluative association (e.g., Elderly = bad). When activated, the model assumes that the association drives responses toward the corresponding key (e.g., toward the key assigned to sort "bad" words).

Detection. This parameter reflects the probability of accurately identifying the stimulus and the task requirements. Here, D estimates the likelihood that the participant correctly perceives the image as belonging to an elderly person and recalls the correct key assignment.

Overcoming bias (OB). This parameter estimates the likelihood of overcoming the association given that both the association (AC) and correct detection (D) have occurred. If the elderly = bad association drives a right-key response, but the correctly detected

[*]Members of higher-status groups do not have to wrestle with this contention, as the hierarchies of society favor their ingroup. For them, internalizing culturally dominant attitudes or ingroup-affirming attitudes produces the same result: favor for their higher-status ingroup over the lower-status outgroup.

[†]Calanchini and colleagues (31) were the first to apply formal process modeling to outgroup bias. However, their outgroup analyses focused entirely on investigating the relative extents of positive versus negative associations, not the extents to which the various cognitive processes distinguished by the model, including self-regulatory control, explained a lack of ingroup preference among lower-status group members.

Downloaded from https://www.pnas.org by COLUMBIA UNIVERSITY on September 9, 2026 from IP address 128.59.177.134.

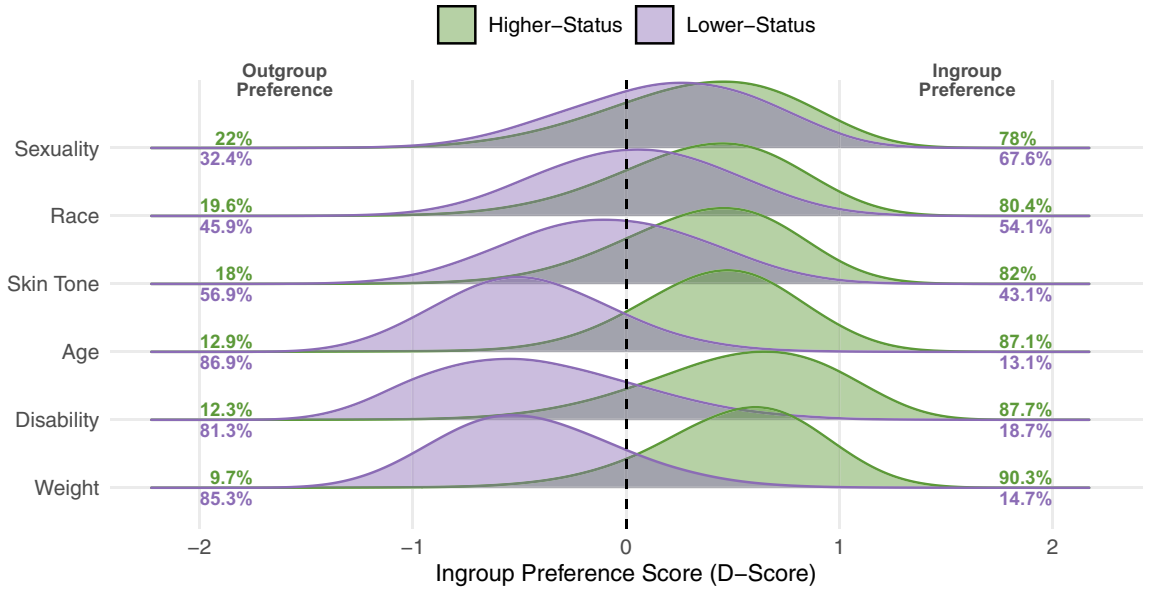

Fig. 1. Distributions reflect IAT D-scores separately for higher-status (green) and lower-status (purple) group members within each attitude domain. Positive D-scores indicate faster responding consistent with ingroup preference; negative D-scores indicate the reverse (i.e., outgroup preference). The dashed vertical line marks zero. Percentages on the *Right* reflect the proportion of respondents within each group status whose D-score was ingroup-affirming; percentages on the *Left* reflect the proportion whose D-score was outgroup-affirming.

Downloaded from https://www.pnas.org by COLUMBIA UNIVERSITY on September 9, 2026 from IP address 128.59.177.134.

task requirement is a left-key response, a conflict arises. The OB parameter estimates the success rate of overcoming the association to resolve this conflict in favor of the correct response.

Guessing. This parameter captures response tendencies when no association is activated and the correct response is not detected. It estimates a bias toward responding with a specific key (e.g., "bad" key vs. "good" key) in the absence of the other processes.

The QPM estimates these parameters based on a set of equations that predict the probability of correct and incorrect responses for each unique trial. To do so, it sums the probabilities of the distinct cognitive paths that lead to a given response (*SI Appendix*, Fig. S1). For instance, on a trial in which the model assumes the activated association conflicts with the correct answer (e.g., a task requiring a response using the "good" key despite a negative association), a correct response can be achieved in three ways. First, the biased association is simply not activated (1 – AC), and the participant correctly detects the answer (D). Second, the bias is activated (AC) and the answer is detected (D), but the biased association is successfully overcome (OB) in favor of the correctly detected answer. Third, if neither the association is activated (1 – AC) nor the answer detected (1 – D), the participant relies on a response bias (G) toward the key that happens to be correct on that trial. Mathematically, the total probability of a correct response on this trial is the sum of these three paths: [(1 – AC) × D] + [AC × D × OB] + [(1 – AC) × (1 – D) × G]. The QPM utilizes maximum likelihood estimation to iteratively adjust the parameter values until the predicted response probabilities align as closely as possible with the observed response frequencies.

The QPM parameters have been extensively validated (38, 39, 48). The present research hinges on interpreting OB as reflecting self-regulatory control, so we briefly review five pieces of converging evidence supporting this psychological interpretation. First, response deadlines reduce OB and D, but not AC and G, consistent with OB reflecting resource-dependent controlled processing (38). Second, alcohol consumption, whose primary cognitive implication is impaired response inhibition (49), selectively impairs OB while leaving other parameters largely intact (39). Third, OB declines with age in a pattern consistent with the broader deterioration of self-regulatory functioning (50). Fourth, intentional strategies to overcome bias selectively increase OB (43). Fifth, neural and response time data indicate that OB reflects the implementation of inhibitory control rather than simply conflict detection or cautious responding (49, 51; see *SI Appendix* for a more detailed review). We therefore use self-regulatory control to refer to the class of processing captured by OB. We remain agnostic about the particular processes by which this regulatory control is enacted.[‡]

Model Specification. To capture the direction of respondents' intergroup bias (ingroup-affirming vs. outgroup-affirming), we specified the QPM with two variants and fit both to each participant's data. We then selected the better-fitting model, per participant, by comparing goodness-of-fit *P*-values (for a similar approach, 31). While the parameters of the QPM are structurally equivalent between the two model specifications, the interpretation of the AC and OB parameters differs:

Ingroup affirming specification. For higher-status respondents (e.g., young), AC encodes the activation of positive associations for their higher-status ingroup (e.g., young–good) and negative associations for the lower-status outgroup (e.g., elderly–bad). For lower-status respondents (e.g., elderly), AC encodes the activation of positive associations for their lower-status ingroup (e.g., elderly–good) and negative associations for the higher-status outgroup (e.g., young-bad). Therefore, OB encodes self-regulatory control over positive evaluations of the ingroup and negative evaluations of the outgroup.

Outgroup affirming specification. For higher-status respondents (e.g., young), AC encodes positive evaluations for the lower-status group (e.g., elderly–good) and negative evaluations for their

[‡]The literature on cognitive control often distinguishes between two classes (41). Proactive control reflects anticipatory, sustained goal maintenance that prevents interference before it arises. Reactive control reflects conflict-triggered inhibition deployed after a biased association has been activated. We do not make claims regarding whether OB reflects proactive control, reactive control, or both, as this question remains empirically open. Nonetheless, the *SI Appendix* offers a detailed review of the evidence bearing on this distinction as it applies to OB.

higher-status ingroup (e.g., young–bad). For lower-status respondents (e.g., elderly), AC encodes positive evaluations for the higher-status outgroup (e.g., young–good) and negative evaluations for their lower-status ingroup (e.g., elderly–bad). Therefore, OB encodes self-regulatory control over positive evaluations of the outgroup and negative evaluations of the ingroup.

The difference in AC and OB estimates between the two variants offers a unique opportunity to test the relative extent of bias activation and control when that bias favors one's ingroup versus outgroup.

Lower-status groups' responses generally align with an outgroup-affirming model. We first tested whether outgroup bias is more evident in lower-status groups relative to higher-status groups, a pattern expected given the IAT behavioral data (Fig. 1). Using chi-square tests of independence across the six attitude domains, we compared the rates at which responses were better fit by an outgroup-affirming (outgroup positivity, ingroup negativity) versus ingroup-affirming QPM. That is, we tested whether the odds of demonstrating outgroup-affirming versus ingroup-affirming bias are greater for lower-status versus higher-status groups.

Across all six attitude domains, lower-status group members showed systematically greater relative odds of their responses being better fit by an outgroup-affirming (vs. ingroup-affirming) model compared to higher-status group members (all $ps < 0.001$). This divergence was most pronounced for the domains of age ($OR = 8.34$), weight ($OR = 7.25$), and disability ($OR = 6.43$), in which lower-status respondents were over six times more likely than higher-status respondents to exhibit outgroup-affirming behavior (Fig. 2). The pattern also held robustly, though to a smaller degree, for skin tone ($OR = 3.02$), race ($OR = 2.18$), and sexuality ($OR = 1.70$; see *SI Appendix*, Tables S4–S10, for summaries of these and all other analyses).

To clarify the nature of these interactions, we tested whether absolute preferences within each group differed from chance (50%) using exact binomial tests. Higher-status respondents' data were consistently fit better by an ingroup-affirming model, across all six domains (all $ps < 0.001$; odds of outgroup-affirming fit: 0.33 to 0.40). Lower-status respondents' behavior, however, bifurcated. For domains of sexuality and race, lower-status respondents' data were best fit by ingroup-affirming models ($ps < 0.001$), indicating the status interaction in these domains reflects attenuated ingroup bias rather than a loss of it. In contrast, for domains of skin tone, age, disability, and weight, lower-status respondents' data were fit better by outgroup-affirming models ($ps < 0.001$). The extent of this outgroup bias varied, ranging from a weak level for skin tone (Odds = 1.08) to moderate levels for disability (Odds = 2.30), weight (Odds = 2.40), and age (Odds = 3.11).

Estimating the Activation of Evaluative Associations and Their Self-Regulation. To estimate cognitive parameters, response frequencies (i.e., correct and incorrect responses per trial type) were summed across respondents within each group status (lower-status, higher-status) × QPM specification (ingroup-affirming, outgroup-affirming) configuration, per IAT. We implemented a joint modeling framework that joins these datasets by model specification, enabling the simultaneous estimation of parameters for both ingroup-affirming and outgroup-affirming specifications within a single model (see *Materials and Methods*). To formally test for parameter differences, we applied equality constraints (e.g., fixing $OB_{ingroup\text{-}affirming} = OB_{outgroup\text{-}affirming}$) and assessed the change in model fit. We concluded that parameters differed significantly if the equality-constrained model produced a significantly worse goodness-of-fit (G^2) compared to the unconstrained model.

Activation of evaluative associations (AC). To test for asymmetries in the activation of evaluative associations, we compared the positive versus negative AC estimates ($\Delta AC = AC_{positive} - AC_{negative}$) for both ingroups and outgroups. If $\Delta AC > 0$, positive (vs. negative) associations of that group are estimated to activate more often, whereas $\Delta AC < 0$ indicates the reverse.

For higher-status groups, the pattern of AC estimates was consistent across all six attitude domains. Among higher-status groups, positive ingroup associations were estimated to activate significantly more often than negative ingroup associations (ΔAC range: 0.03 to 0.06) but negative outgroup associations were estimated to activate significantly more often than positive outgroup associations (ΔAC range: −0.01 to −0.05; $ps < 0.001$; see Fig. 3). Overall, among higher-status groups, ingroup-affirming associations were activated more frequently than outgroup-affirming associations.

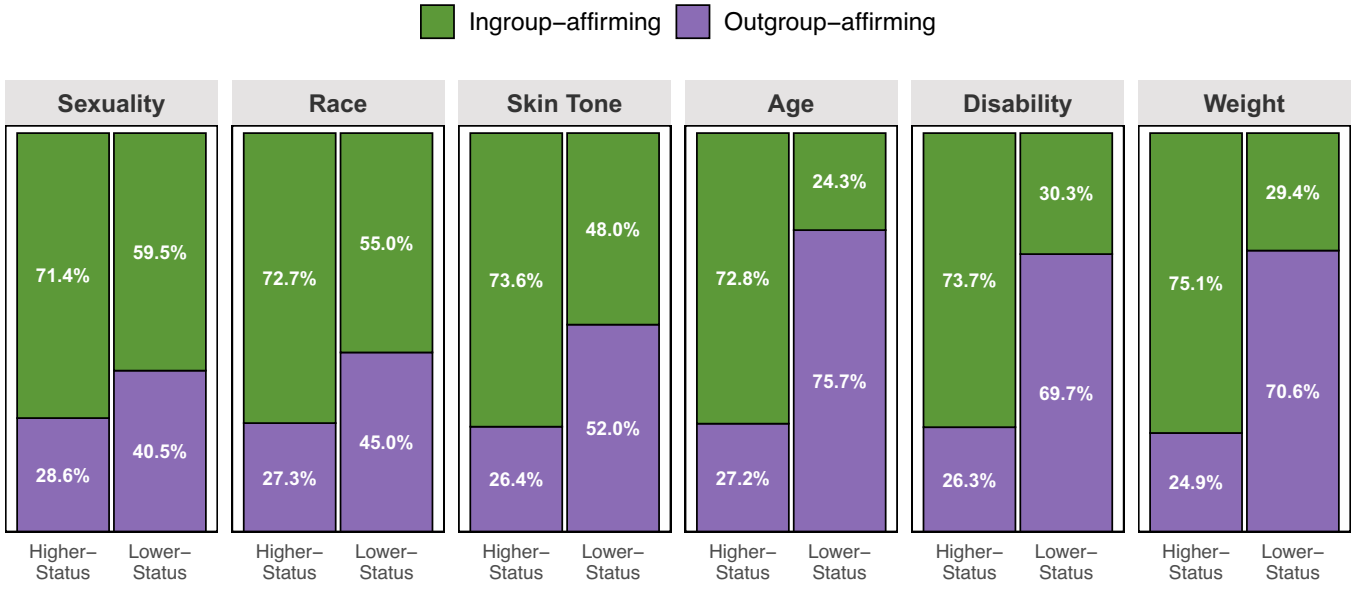

Fig. 2. Bars reflect the percentage of respondents within each IAT and group status combination whose data were best fit by one of two QPM specifications. Panel columns represent social category domains. The x axis distinguishes whether the bars reflect higher-status or lower-status respondents. Bar colors denote whether the data were best fit by a model that assumes positive evaluations of ingroups and negative evaluations of outgroups (i.e., ingroup-affirming; green) or negative evaluations of ingroups and positive evaluations of outgroups (i.e., outgroup-affirming; purple).

Downloaded from https://www.pnas.org by COLUMBIA UNIVERSITY on September 9, 2026 from IP address 128.59.177.134.

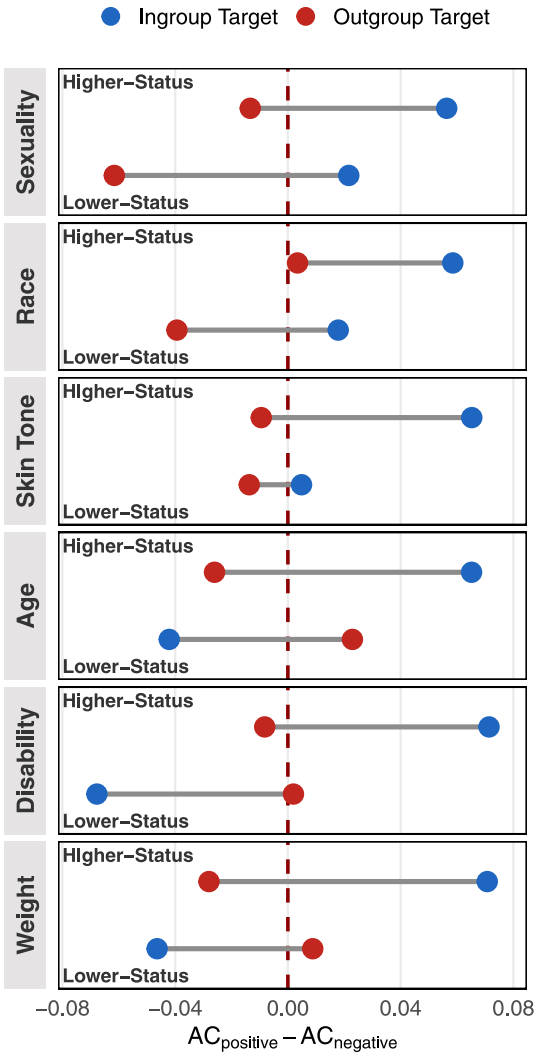

Fig. 3. Error bars represent 95% CI (note that these CIs are very small due to large sample sizes and thus are obscured by the point marker). The x-axis displays the difference between the estimated probability of activating positive versus negative associations ($AC_{positive}$ – $AC_{negative}$). Positive values (to the *Right* of the dashed red line) indicate that positive associations are more likely to be activated than negative associations. Negative values (to the *Left* of the dashed red line) indicate the opposite. Dot colors denote whether the estimates correspond to Ingroup targets (blue) or Outgroup targets (red). Panel rows represent attitude domains. The y-axis represents respondents' group status, with the *Top* row displaying higher-status groups and the *Bottom* row displaying lower-status groups.

For lower-status groups, the pattern was less consistent. For the domains of sexuality, race, and skin tone, positive ingroup associations were estimated to activate significantly more often than negative ingroup associations (ΔAC = 0.08, 0.03, and 0.01, respectively) and negative outgroup associations were estimated to activate significantly more often than positive outgroup associations (ΔAC = –0.01, –0.03, and –0.01, respectively; *p*s < 0.001). However, among lower-status groups in the domains of age, disability, and weight, positive outgroup associations were estimated to activate significantly more often than negative outgroup associations (ΔAC = 0.04, 0.05, and 0.03, respectively), and negative ingroup associations were estimated to activate significantly more often than positive ingroup associations (ΔAC = –0.03, –0.02, and –0.03, respectively; *p*s < 0.001). Overall, among lower-status groups, ingroup-affirming associations were more frequently activated for the domains of sexuality, race, and skin tone but outgroup-affirming associations were more frequently activated for the domains of age, disability, and weight.

OB. To test whether self-regulatory control plays a role in the relative absence of ingroup bias among lower-status versus higher-status groups, we compared OB estimates as a function of group status and QPM specification for each of the six attitude domains.

For higher-status group members, the pattern of $OB_{ingroup\text{-}affirming}$ versus $OB_{outgroup\text{-}affirming}$ was inconsistent (Fig. 4; *Top* row panels). In the domains of race (ΔOB = –0.29), skin tone (ΔOB = –0.07), and weight (ΔOB = –0.03), higher-status respondents were estimated to overcome outgroup-affirming evaluations significantly more than ingroup-affirming evaluations (*p*s < 0.001). The opposite pattern was found for the domains of age (ΔOB = 0.23) and sexuality (ΔOB = 0.12; *p*s < 0.001). For the disability IAT, no significant difference emerged between $OB_{ingroup\text{-}affirming}$ and $OB_{outgroup\text{-}affirming}$ (ΔOB = 0.00, *p* = 1.00). Overall, differences in OB were inconsistent among higher-status groups.

For lower-status groups, the pattern of OB was highly consistent (Fig. 4; *Bottom* row panels). Across five of the six attitude domains, ingroup-affirming evaluations were estimated to be overcome significantly more often than outgroup-affirming evaluations. The OB estimates suggest that positive evaluations of the lower-status ingroup and negative evaluations of the higher-status outgroup were more likely to be overcome than negative evaluations of the ingroup and positive evaluations of the outgroup (*p*s < 0.001), with the exception of the disability domain (*pp* = 0.247), which descriptively showed the same pattern. The effect was strongest for the domains of sexuality (ΔOB = 0.67), race (ΔOB = 0.67), and skin tone (ΔOB = 0.47), followed by weight (ΔOB = 0.46), age (ΔOB = 0.32), and disability (ΔOB = 0.02). Overall, among lower-status groups, ingroup-affirming evaluations were overcome more often than outgroup-affirming evaluations, although the magnitude of this difference varied across attitudinal domains.

Differences in overcoming positive and negative evaluations. Thus far, our results suggest that, among lower-status groups, ingroup-affirming evaluations (i.e., ingroup-positive/outgroup-negative) are overcome more often than outgroup-affirming evaluations (i.e., ingroup-negative/outgroup-positive). But, as noted, these ingroup-affirming evaluations reflect two distinct components: positive evaluations of the ingroup and negative evaluations of the outgroup. Thus, it remains an open question whether the overcoming of ingroup-affirming evaluations is characterized more by overcoming positive ingroup evaluations (e.g., pro-gay) or negative outgroup evaluations (e.g., anti-straight). To address this, we refit the QPM, separating OB into the overcoming of positive evaluations and the overcoming of negative evaluations. For example, for the domain of age, the ingroup-affirming model distinguished between overcoming positive evaluations of elderly people and negative evaluations of young people.

In four of the six domains, OB estimates for ingroup-affirming evaluations reflected greater overcoming of positive (lower-status) ingroup evaluations (Fig. 5, *Left*-most blue bars in each panel) than negative (higher-status) outgroup evaluations (Fig. 5, *Left*-most red bars in each panel; ΔOB range = 0.06 to 0.41): such effects were strongest in domains of weight (ΔOB = 0.41), followed by disability (ΔOB = 0.29), sexuality (ΔOB = 0.27), and race (ΔOB = 0.06; *p*s < 0.001). Conversely, for the domains of age and skin tone, OB estimates for ingroup-affirming evaluations reflected relatively greater overcoming of outgroup-negative than ingroup-positive evaluations (ΔOBs = –0.49 and –0.11, respectively; *p*s < 0.001). Estimates for the overcoming of outgroup-affirming evaluations (right-most sets of bars) showed less consistent patterns and lower overall estimates (*SI Appendix*).

Downloaded from https://www.pnas.org by COLUMBIA UNIVERSITY on September 9, 2026 from IP address 128.59.177.134.

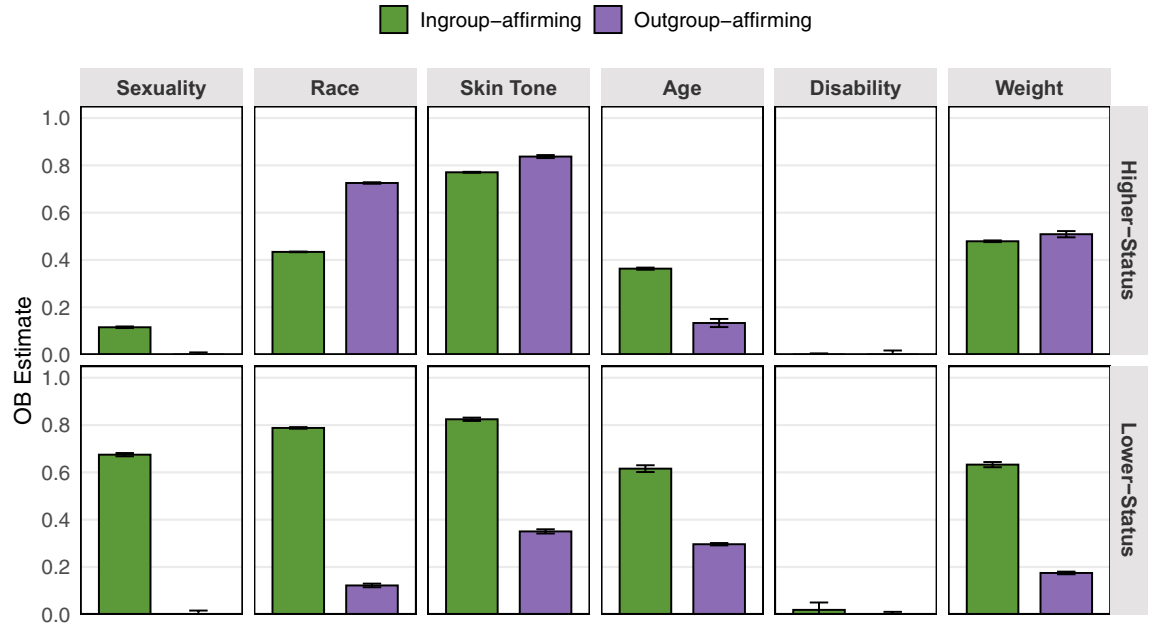

Fig. 4. Error bars represent 95% CI. Bars reflect the OB parameter estimates. Values of OB < 0.50 indicate that OB occurs on less than half of relevant trials, whereas OB > 0.50 indicates the opposite. Bar colors denote whether the estimates come from a model that assumes the overcoming of positive evaluations of ingroups and negative evaluations of outgroups (i.e., ingroup-affirming; green) or the overcoming of negative evaluations of ingroups and positive evaluations of outgroups (i.e., outgroup-affirming; purple). Panel columns represent attitude domains. Panel rows represent respondents' group status, with the *Top* row displaying higher-status groups and the *Bottom* row displaying lower-status groups.

Discussion

The present research implicates self-regulatory control in the absence of implicit ingroup preference among lower-status groups. Associative learning provides one plausible pathway through which outgroup-affirming evaluations could be acquired (29, 30, 32). Our estimates for the activation of evaluative associations (AC) are broadly consistent with such accounts. But the present results suggest that the absence of ingroup preference emerges not only at the level of associative processing but also at the level of regulatory control, with ingroup-affirming evaluations accompanied by greater self-regulatory control than outgroup-affirming evaluations among lower-status groups. Notably, the asymmetry in self-regulatory control was broadly consistent across attitude domains, including those in which lower-status groups showed only attenuated ingroup preference rather than outright outgroup preference (e.g., sexuality). These results therefore identify self-regulatory control as a previously unexplored process that contributes across attitude domains toward an absence of ingroup preference among lower-status groups.

What Drives the Asymmetry in Self-Regulatory Control? The finding that, among lower-status groups, ingroup-affirming evaluations are regulated more often than outgroup-affirming evaluations raises the question of how such an asymmetry is acquired. One possibility we consider centers on the social environment itself. Prior research demonstrates that counterstereotypical behavior among lower-status group members draws social penalties that function

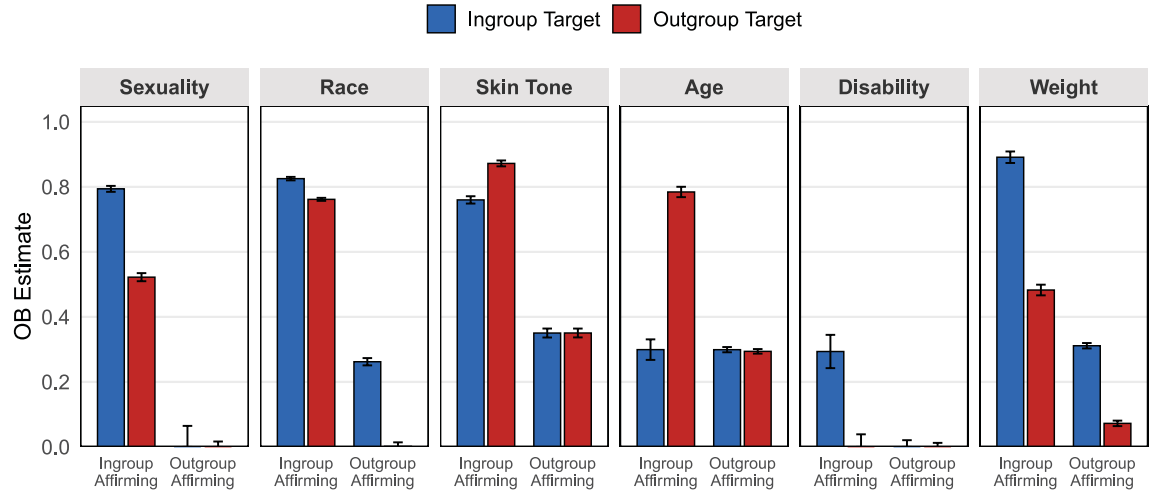

Fig. 5. Error bars represent 95% CI. Bars reflect the OB parameter estimates. Values of OB < 0.50 indicate that OB successfully occurs on less than half of relevant trials, whereas OB > 0.50 indicates OB successfully occurs on more than half of relevant trials. Bar colors denote whether OB estimates the overcoming of ingroup evaluations (positive for the ingroup-affirming model; negative for the outgroup-affirming model), or outgroup evaluations (negative for the ingroup-affirming model; positive for the outgroup-affirming model). Panel columns represent attitude domains.

Downloaded from https://www.pnas.org by COLUMBIA UNIVERSITY on September 9, 2026 from IP address 128.59.177.134.

to maintain existing hierarchies (52, 53). These penalties may promote self-regulatory control over ingroup-affirming behavior. For example, high-performing Black Americans report suppressing displays of academic success to avoid backlash, and sexual minorities report chronic pressure toward identity concealment to avoid prejudice and discrimination (53, 54). This perspective resonates with the view that implicit bias reflects the statistical regularities of the social environment (28, 55): just as society provides signals that shape evaluations, it also may convey signals that elicit self-regulatory control in line with hierarchy maintenance. The OB asymmetry found here may thus reflect a regulatory response to chronic environmental pressure to withhold ingroup-affirming behavior. Whether and how such environmental signals give rise to the specific regulatory asymmetry observed on the IAT remains an open question that future research should directly examine.

These data raise several empirical questions worth pursuing. First, what drives variability in the magnitude of this regulatory asymmetry across attitude domains, and why does its magnitude not closely track that of behavioral outgroup bias? The asymmetry in OB was most pronounced for sexuality and race domains and less pronounced for weight, age, and disability domains. This pattern warrants investigation into whether domain-level features (e.g., differences in the frequency or severity of social penalties for lower-status ingroup-affirmation) predict the magnitude of the effect. Relatedly, the two-OB decomposition indicates that the regulated content also varied: in age and skin tone, the asymmetry reflected greater control over negative evaluations of the higher-status outgroup than positive evaluations of the lower-status ingroup.

Notably, lower-status disability group members showed no significant difference in overcoming ingroup-affirming versus outgroup-affirming evaluations, with near-zero aggregate-level OB estimates across QPM specifications. The pattern ($OB_{ingroup\text{-}affirming} > OB_{outgroup\text{-}affirming}$) was nevertheless descriptively present in the aggregate analyses, and also emerged as a significant effect in the two-OB model extension, as well as in a supplemental analysis using individual-level data (*SI Appendix*), albeit with lower estimates than in other domains. Additionally, because aggregate-level estimation is susceptible to boundary compression when true OB values are small (56), the pattern in OB across the other five attitude domains cannot be entirely ruled out in the disability domain, but overall disability OB estimates are clearly smaller than in other domains. Future research should explore why, among lower-status groups, ingroup-affirming evaluations are disproportionately overcome more in some domains than others.

Second, what are the longer-term consequences of chronically preventing ingroup-affirming evaluations? Research on inhibition-induced devaluation suggests that repeated inhibition of responses to a stimulus can reduce its perceived positivity (42, 44, 57, 58), raising the question of whether such regulatory asymmetries gradually erode ingroup-affirming associations over time. If so, a consequence may ultimately be a shift in the balance of how the ingroup is evaluated relative to the outgroup, as revealed, for example, through the AC parameter. Longitudinal designs tracking both AC and OB parameters alongside individual differences in social experience would be well-positioned to examine these questions. If ingroup-affirming evaluations are chronically accompanied by regulatory control among lower-status groups, such repeatedly enacted responses may also, over time, transition from effortful and intentional to relatively automatic (45).

The Role of Motivation. According to System Justification Theory, lower-status group members with a greater desire to justify the societal status quo are less likely to demonstrate ingroup preference (18). However, the cognitive implementation of this motivation remains unclear. Does the desire to see the system as legitimate simply lead to the acquisition of outgroup-affirming associations, or does it actively recruit the regulatory processes identified in the current research, or both? It is possible that while the social environment may condition self-regulatory control over lower-status groups' ingroup-affirming behaviors, motivation to justify the system may potentiate this effect, further reducing the likelihood of ingroup affirmation. Future research should examine whether system-justifying motives predict self-regulatory control over ingroup-affirming evaluations among lower-status groups, the prevalence of those evaluations, or both, clarifying how individual psychological needs support social hierarchy maintenance.

Inconsistent Self-Regulatory Control Among Higher-Status Groups. Whereas ingroup-affirming evaluations were often accompanied by greater self-regulatory control than outgroup-affirming evaluations among lower-status groups, no consistent pattern emerged among higher-status groups. Extensive literature on the psychological implications of power suggests that individuals with more power generally feel freer to act on their impulses (59, 60). Groups who sit at the top of the social hierarchy may be subject to fewer systemic pressures on self-regulatory control in any particular direction. This may explain why estimates of higher-status self-regulatory control lack directional uniformity across domains. What accounts for the specific patterns observed within each domain, however, remains an open question for future research.

Limitations. The current research utilizes a dataset of unprecedented scale for cognitive modeling of intergroup behavior, but it is not without limitations. Primary among them is the estimation approach. The main analyses relied on aggregate-level estimates using maximum likelihood estimation, which assumes independently and identically distributed data and results in complete pooling of information across respondents. This allowed us to leverage the full statistical power of the dataset, but did not account for individual-level variability. A no-pooling approach relying on individual QPM estimates was considered; however, because IAT data typically contain few errors, many individual-level OB estimates failed to converge or had SEs covering nearly the entire probability space, requiring the exclusion of ~80% of the sample. The remaining individual-level estimates produced the same pattern of results as the aggregate-level analyses (*SI Appendix*). Furthermore, MPT parameters converge well across estimation approaches (56). One resolution to this trade-off is the partial-pooling of information through hierarchical Bayesian estimation. Though computationally prohibitive for datasets of this magnitude, this approach would stabilize individual estimates by leveraging group-level distributions.

A second limitation concerns the operationalization of group status for skin tone and age. For skin tone, race served as a proxy for a continuous construct, which may mask some variation in colorism-related evaluation. For age, the cutoffs used to classify respondents are necessarily somewhat arbitrary, leaving unanswered what the more precise boundaries are. Future work should examine finer-grained operationalizations of both constructs.

Conclusion. For decades, the absence of ingroup preference among lower-status groups has been understood primarily as a reflection of internalized stigma. By applying formal process modeling to 5.7 million implicit bias tests and six social group categories, we demonstrate that this phenomenon is not merely a product of learned associations but also of the regulatory control

Downloaded from https://www.pnas.org by COLUMBIA UNIVERSITY on September 9, 2026 from IP address 128.59.177.134.

that accompanies them. Among lower-status groups, ingroup-affirming evaluations are accompanied by greater self-regulatory control than outgroup-affirming evaluations. Understanding how social hierarchies are sustained may therefore require looking beyond the content of our biases to the cognitive machinery regulating their expression.

Materials and Methods

Participants. Data were retrieved from N = 6,119,315 volunteer respondents to the Project Implicit demonstration website (https://implicit.harvard.edu/), who provided informed consent and demographic information before completing an IAT related to age, disability, race, sexuality, skin tone, or weight. All data were deidentified. The six domains were selected for two primary reasons. First, they span a theoretically meaningful range of stigma types from sociodemographic (race, sexuality) to body-related (age, disability, body weight), with skin tone conceptually situated between these classes. This distinction is associated with variation in intergroup evaluation among disadvantaged group members (61, 62), thereby permitting an assessment of whether the present findings generalize across qualitatively distinct attitude domains. Second, the data were highly accessible given their prior use by the authors.

Data were only included if individual datasets had complete IAT data (i.e., 120 critical trials). Additionally, respondents were only included if they self-reported a group status relevant to the specific IAT completed (e.g., reporting as Black or White for the Race IAT). Data inclusion began January 1, 2007, and ended May 31, 2022, for about 16 years of data. For 454,622 individual datasets (7.4%), neither QPM specification fit the data better than the other and were therefore excluded from further analyses. The final sample comprised 5,664,693 respondents (see *SI Appendix*, Table S1, for demographics and sample sizes).

Materials and Procedure. The IAT (46) is a dual-categorization task designed to assess the association people have between certain groups and attributes. When taking an evaluative IAT, participants are instructed to sort groups (e.g., Black vs. White people) and evaluative words (e.g., "wonderful" vs. "evil") using one of two keys. Throughout the task, correctly categorizing positive (negative) stimuli always requires the same key response. The correct key response for each group, however, varies across trial blocks, such that the positive (negative) key sometimes is shared with one group (e.g., Black people) and sometimes is shared with the other group (e.g., White people).

Group status for each attitude domain was determined using respondents' self-reported demographic information. For the race IAT, White-identifying respondents were classified as higher-status and Black-identifying respondents as lower-status. For the skin tone IAT, which lacks a directly corresponding demographic identity, racial self-identification was used as a proxy, applying the same White/Black classification. For the age IAT, respondents aged 20 and under were classified as higher-status and those aged 40 and over as lower-status, consistent with cutoffs used in prior large-scale analyses of Project Implicit data determined by the youngest and oldest age brackets with sufficient sample size (63). For the sexuality, disability, and weight IATs, the relevant demographic question was administered only to respondents completing that specific IAT. Straight-identifying respondents were classified as higher-status and gay- or lesbian-identifying respondents as lower-status. Nondisabled respondents were classified as higher-status and those reporting a disability as lower-status. For body weight, respondents were classified using a 7-point scale ranging from very underweight to very overweight; those at or below the midpoint were classified as higher-status and those above the midpoint as lower-status.

Analytic Approach. All QPM fitting was conducted with maximum likelihood estimation (MLE) using the MPTinR package (64). Response frequencies (i.e., counts of correct and incorrect responses for each of eight unique trial types; see *SI Appendix* for a detailed review) were summed across all respondents within each Group Status × Model Specification cell, per IAT, prior to model fitting. This aggregation results in a single set of parameter estimates per cell, rendering standard General Linear Model analyses impossible due to a lack of variance. Consequently, hypothesis testing relied on the likelihood-ratio test statistic G^2, which follows a χ^2 distribution (64, 65). Parameter estimates were compared by computing the difference in goodness-of-fit (ΔG^2) of unconstrained models against nested models with equality constraints imposed on the relevant parameters, as described in the *Results*. A significant reduction in model fit (ΔG^2) indicated that the parameters differed significantly.

Data, Materials, and Software Availability. Anonymized de-identified behavioral data, model files and analysis code data have been deposited in Open Science Framework (OSF) (https://osf.io/83epy/overview) (66). All other study data are included in the article and/or supporting information.

Author affiliations: [a]Department of Psychology, Columbia University, New York, NY 10027; [b]Kellogg School of Management, Northwestern University, Evanston, IL 60208; [c]Department of Psychology, University of Marburg, Marburg 35032, Germany; [d]Department of Psychology, Harvard University, Cambridge, MA 02138; and [e]Department of Psychology, University of California, Davis, CA 95616

Author contributions: S.A.W.K., T.E.S.C., M.R.B., and J.W.S. designed research; S.A.W.K. performed research; S.A.W.K. and D.W.H. contributed new reagents/analytic tools; S.A.W.K. analyzed data; S.A.W.K. and T.E.S.C. writing - Review & Editing; Validation of results and methods; Data curation and preprocessing; D.W.H. writing - Review & Editing; Validation of results and methods; M.R.B. and J.W.S. writing - Review & Editing; and S.A.W.K. wrote the paper.

1. R. Bergh, N. Akrami, "Generalized prejudice: Old wisdom and new perspectives" in *The Cambridge handbook of the psychology of prejudice*, C. G. Sibley, F. K. Barlow, Eds. (Cambridge University Press, 2017), pp. 438–460.
2. D. Balliet, J. Wu, C. K. W. De Dreu, Ingroup favoritism in cooperation: A meta-analysis. *Psychol. Bull.* **140**, 1556–1581 (2014).
3. L. Goette, D. Huffman, S. Meier, The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *Am. Econ. Rev.* **96**, 212–216 (2006).
4. M. Foddy, M. J. Platow, T. Yamagishi, Group-based trust in strangers: The role of stereotypes and expectations. *Psychol. Sci.* **20**, 419–422 (2009).
5. M. Levine, A. Prosser, D. Evans, S. Reicher, Identity and emergency intervention: How social group membership and inclusiveness of group boundaries shape helping behavior. *Pers. Soc. Psychol. Bull.* **31**, 443–453 (2005).
6. K. Hugenberg, S. G. Young, M. J. Bernstein, D. F. Sacco, The categorization-individuation model: An integrative account of the other-race recognition deficit. *Psychol. Rev.* **117**, 1168–1187 (2010).
7. J. J. Van Bavel, D. J. Packer, W. A. Cunningham, The neural substrates of in-group bias: A functional magnetic resonance imaging investigation. *Psychol. Sci.* **19**, 1131–1139 (2008).
8. M. Cikara, M. M. Botvinick, S. T. Fiske, Us versus them: Social identity shapes neural responses to intergroup competition and harm. *Psychol. Sci.* **22**, 306–313 (2011).
9. X. Xu, X. Zuo, X. Wang, S. Han, Do you feel my pain? Racial group membership modulates empathic neural responses. *J. Neurosci.* **29**, 8525–8529 (2009).
10. H. Bernhard, E. Fehr, U. Fischbacher, Parochial altruism in humans. *Nature* **442**, 912–915 (2006).
11. J. J. Jordan, M. Hoffman, P. Bloom, D. G. Rand, Third-party punishment as a costly signal of trustworthiness. *Nature* **530**, 473–476 (2016).
12. S. Otten, The minimal group paradigm and its maximal impact in research on social categorization. *Curr. Opin. Psychol.* **11**, 85–89 (2016).
13. M. B. Brewer, The psychology of prejudice: Ingroup love and outgroup hate? *J. Soc. Issues* **55**, 429–444 (1999).
14. F. Fu *et al.*, Evolution of in-group favoritism. *Sci. Rep.* **2**, 460 (2012).
15. K. B. Clark, M. P. Clark, Racial identification and preference in Negro children. *Read. Social Psycholo.* 169–178 (1947).
16. J. T. Jost, M. R. Banaji, The role of stereotyping in system-justification and the production of false consciousness. *Br. J. Soc. Psychol.* **33**, 1–27 (1994).
17. J. R. Axt, T. Moran, Y. Bar-Anan, Simultaneous ingroup and outgroup favoritism in implicit social cognition. *J. Exp. Soc. Psychol.* **79**, 275–289 (2018).
18. J. T. Jost, M. R. Banaji, B. A. Nosek, A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Polit. Psychol.* **25**, 881–919 (2004).
19. L. A. Rudman, J. Feinberg, K. Fairchild, Minority members' implicit attitudes: Automatic ingroup bias as a function of group status. *Soc. Cogn.* **20**, 294–320 (2002).
20. L. Ashburn-Nardo, M. L. Knowles, M. J. Monteith, Black Americans' implicit racial associations and their implications for intergroup judgment. *Soc. Cogn.* **21**, 61–87 (2003).
21. R. W. Livingston, The role of perceived negativity in the moderation of African Americans' implicit and explicit racial attitudes. *J. Exp. Soc. Psychol.* **38**, 405–413 (2002).
22. D. S. March, R. Graham, Exploring implicit ingroup and outgroup bias toward Hispanics. *Group Process. Intergroup Relat.* **18**, 89–103 (2015).
23. A.-K. Newheiser, Y. Dunham, A. Merrill, L. Hoosain, K. R. Olson, Preference for high status predicts implicit outgroup bias among children from low-status groups. *Dev. Psychol.* **50**, 1081–1090 (2014).
24. H. Tajfel, *Human Groups and Social Categories* (Cambridge University Press, New York, 1981).
25. H. Tajfel, J. C. Turner, "An integrative theory of intergroup conflict" in *The Social Psychology of Intergroup Relations* (Brooks/Cole, 1979), pp. 33–47.
26. J. T. Jost, B. W. Pelham, O. Sheldon, B. N. Sullivan, Social inequality and the reduction of ideological dissonance on behalf of the system: Evidence of enhanced system justification among the disadvantaged. *Eur. J. Soc. Psychol.* **33**, 13–36 (2003).
27. M. B. Oliver, "Race and crime in the media: Research from a media effects perspective" in *A Companion to Media Studies* (2003), pp. 421–436.
28. H. A. Vuletich, B. K. Payne, Stability and change in implicit bias. *Psychol. Sci.* **30**, 854–862 (2019).

Downloaded from https://www.pnas.org by COLUMBIA UNIVERSITY on September 9, 2026 from IP address 128.59.177.134.

29. R. H. Fazio, "Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework" in *Advances in Experimental Social Psychology*, (Academic Press, 1990), vol. 23, pp. 75–109.
30. B. Gawronski, G. V. Bodenhausen, Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychol. Bull.* **132**, 692 (2006).
31. J. Calanchini, K. Schmidt, J. W. Sherman, S. A. Klein, The contributions of positive outgroup and negative ingroup evaluation to implicit bias favoring outgroups. *Proc. Nat. Acad. Sci.* **119**, e2116924119 (2022).
32. D. S. March, L. Gaertner, M. A. Olson, On the prioritized processing of threat in a dual implicit process model of evaluation. *Psychol. Inq.* **29**, 1–13 (2018).
33. D. S. March, Perceiving a danger within: Black Americans associate Black men with physical threat. *Soc. Psychol. Personal. Sci.* **14**, 942–951 (2023).
34. B. Gawronski, A. Hahn, "Implicit measures: Procedures, use, and interpretation" in *Measurement in Social Psychology*, (Routledge, 2018), pp. 29–55.
35. S. A. Klein, J. W. Sherman, "Measuring and modeling implicit cognition" in *The Routledge Handbook of Philosophy and Implicit Cognition*, (Routledge, 2022), pp. 44–55.
36. J. W. Sherman, K. C. Klauer, T. J. Allen, "Mathematical modeling of implicit social cognition: The machine in the ghost" in *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications*, B. Gawronski, B. K. Payne, Eds. (Guilford Press, New York, 2010), pp. 156–175.
37. J. W. Sherman, B. Gawronski, Y. Trope, Eds., *Dual-process Theories of The Social Mind* (Guilford Press, New York, NY, 2014).
38. F. R. Conrey, J. W. Sherman, B. Gawronski, K. Hugenberg, C. J. Groom, Separating multiple processes in implicit social cognition: The Quadruple Process model of implicit attitudes. *J. Personality Social Psychol.* **89**, 469–487 (2005).
39. J. W. Sherman *et al.*, The self-regulation of automatic associations and behavioral impulses. *Psychol. Rev.* **115**, 314 (2008).
40. D. M. Amodio, J. K. Swencionis, Proactive control of implicit bias: A theoretical model and implications for behavior change. *J. Pers. Soc. Psychol.* **115**, 255 (2018).
41. T. S. Braver, The variable nature of cognitive control: A dual mechanisms framework. *Trends Cogn. Sci.* **16**, 106–113 (2012).
42. V. Allom, B. Mullan, M. Hagger, Does inhibitory control training improve health behaviour? A meta-analysis. *Health Psychol. Rev.* **10**, 168–186 (2016).
43. J. Calanchini, C. K. Lai, K. C. Klauer, Reducing implicit racial preferences: III. A process-level examination of changes in implicit preferences. *J. Pers. Soc. Psychol.* **121**, 796 (2021).
44. Z. Chen, H. Veling, A. Dijksterhuis, R. W. Holland, How does not responding to appetitive stimuli cause devaluation: Evaluative conditioning or response inhibition? *J. Exp. Psychol. Gen.* **145**, 1687 (2016).
45. L. Spierer, C. F. Chavan, A. L. Manuel, Training-induced behavioral and brain plasticity in inhibitory control. *Front. Hum. Neurosci.* **7**, 427 (2013).
46. A. G. Greenwald, D. E. McGhee, J. L. K. Schwartz, Measuring individual differences in implicit cognition: The implicit association test. *J. Pers. Soc. Psychol.* **74**, 1464–1480 (1998).
47. A. G. Greenwald, B. A. Nosek, M. R. Banaji, Understanding and using the implicit association test: I. An improved scoring algorithm. *J. Pers. Soc. Psychol.* **85**, 197 (2003).
48. J. Calanchini, J. W. Sherman, Implicit attitudes reflect associative, non-associative, and non-attitudinal processes. *Soc. Personal. Psychol. Compass* **7**, 654–667 (2013).
49. B. D. Bartholow, A. Heinz, Alcohol and aggression without consumption: Alcohol cues, aggressive thoughts, and hostile perception bias. *Psychol. Sci.* **17**, 30–37 (2006).
50. K. Gonsalkorale, J. W. Sherman, K. C. Klauer, Aging and prejudice: Diminished regulation of automatic race bias among older adults. *J. Exp. Soc. Psychol.* **45**, 410–414 (2009).
51. K. Gonsalkorale, J. W. Sherman, K. C. Klauer, Measures of implicit attitudes may conceal differences in implicit associations. *Soc. Psychol. Personal. Sci.* **5**, 271–278 (2013).
52. L. A. Rudman, K. Fairchild, Reactions to counterstereotypic behavior: The role of backlash in cultural stereotype maintenance. *J. Pers. Soc. Psychol.* **87**, 157–176 (2004).
53. J. E. Phelan, L. A. Rudman, Reactions to ethnic deviance: The role of backlash in racial stereotype maintenance. *J. Pers. Soc. Psychol.* **99**, 265 (2010).
54. I. H. Meyer, Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: Conceptual issues and research evidence. *Psychol. Bull.* **129**, 674 (2003).
55. B. K. Payne, H. A. Vuletich, K. B. Lundberg, The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychol. Inq.* **28**, 233–248 (2017).
56. H. Singmann *et al.*, Evaluating the robustness of parameter estimates in cognitive models: A meta-analytic review of multinomial processing tree models across the multiverse of estimation methods. *Psychol. Bull.* **150**, 965 (2024).
57. A. Jones *et al.*, Inhibitory control training for appetitive behaviour change: A meta-analytic investigation of mechanisms of action and moderators of effectiveness. *Appetite* **97**, 16–28 (2016).
58. H. Veling, R. W. Holland, A. van Knippenberg, When approach motivation and behavioral inhibition collide: Behavior regulation through stimulus devaluation. *J. Exp. Soc. Psychol.* **44**, 1013–1019 (2008).
59. A. D. Galinsky, D. H. Gruenfeld, J. C. Magee, From power to action. *J. Pers. Soc. Psychol.* **85**, 453–466 (2003).
60. D. Keltner, D. H. Gruenfeld, C. Anderson, Power, approach, and inhibition. *Psychol. Rev.* **110**, 265–284 (2003).
61. T. E. Charlesworth, N. Sanjeev, M. L. Hatzenbuehler, M. R. Banaji, Identifying and predicting stereotype change in large language corpora: 72 groups, 115 years (1900–2015), and four text sources. *J. Pers. Soc. Psychol.* **125**, 969 (2023).
62. I. Essien, J. Calanchini, J. Degner, Moderators of intergroup evaluation in disadvantaged groups: A comprehensive test of predictions from system justification theory. *J. Pers. Soc. Psychol.* **120**, 1204 (2021).
63. T. E. S. Charlesworth, M. R. Banaji, Patterns of implicit and explicit attitudes: IV. Change and stability from 2007 to 2020. *Psychol. Sci.* **33**, 1347–1371 (2022).
64. H. Singmann, D. Kellen, MPTinR: Analysis of multinomial processing tree models in R. *Behav. Res. Methods* **45**, 560–575 (2013).
65. D. M. Riefer, W. H. Batchelder, Multinomial modeling and the measurement of cognitive processes. *Psychol. Rev.* **95**, 318–339 (1988).
66. S. A. W. Klein, T. E. S. Charlesworth, D. W. Heck, M. R. Banaji, J. W. Sherman, Data and analysis code for "A role of self-regulatory control in the absence of ingroup preference." Open Science Framework. https://osf.io/83epy/overview?view_only=ff82509f7516462db58063555bb3609a. Deposited 27 December 2025.

Downloaded from https://www.pnas.org by COLUMBIA UNIVERSITY on September 9, 2026 from IP address 128.59.177.134.