Supporting Information

A Role for Self-Regulatory Control in the Absence of Ingroup Preference

Public repository of data and analysis code: https://osf.io/83epy

Table of Contents

Demographics ……. 3

Behavioral Data: The IAT D-Score ……4

The Quadruple Process Model …… 7

Ingroup-Affirming vs. Outgroup-Affirming Specifications …….. 8

Model Specification and Comparison ……10

Parameter Analyses ……13

Parameter Estimates …….. 19

Results of the Detection (D) and Guessing (G) Parameters …….. 28

Detection (D) ……..28

Guessing (G) ……..29

Model Fit ……. 30

Quadruple Process Model with Two OB Parameters ……... 32

Individual-Level Parameter Estimation ……... 37

Methodological Considerations and Data Exclusions ……..37

Individual-Level OB Analyses ……... 39

Age-Based Robustness Analyses ……... 46

Overcoming Bias as a Measure of Self-Regulatory Control……... 54

Evidence for OB as a Measure of Self-Regulatory Control ……..... 54

Distinguishing Types of Self-Regulatory Control: Proactive Versus Reactive Control 57

The Quad Model Equations …….. 60

Example QPM Equations ..……. 61

References …… 62

Demographics

Table S1

Sample Demographics by IAT Domain

IAT	*N*	Age *M* (*SD*)	% White	% Female	% Lower-Status
Sexuality	907,360	27.08 (12.11)	78.8%	64.8%	11.8%
Race	2,776,016	30.32 (13.72)	86.5%	63.2%	13.5%
Skin Tone	657,351	30.82 (13.31)	80.2%	70.3%	13.5%
Age	478,258	31.69 (17.87)	80.1%	68.7%	40.9%
Disability	343,332	29.62 (12.82)	82.0%	74.2%	17.5%
Weight	502,376	28.75 (12.61)	79.9%	71.6%	44.5%

Note. Sample characteristics for respondents included in the final analyses. Respondents were included if they self-reported a group status relevant to the specific IAT completed and had complete IAT data.

Behavioral Data: The IAT D-Score

The D-score and the QPM parameters are derived from fundamentally different aspects of IAT performance. The D-score is computed from response latencies: mean latencies are computed separately for each critical block, their difference is taken, and the result is divided by the pooled standard deviation of response times across both blocks (1). For the current research, D-scores are computed such that positive scores always reflect preferential responding toward the ingroup (i.e., faster responses when the respondents' ingroup is paired with positive evaluative words) and negative scores always reflect preferential responding toward the outgroup.

Higher-status group members showed reliably positive average IAT D-scores across all six attitude domains (Table S2), meaning they were consistently faster to respond when their ingroup was paired with positive evaluative words (i.e., ingroup preference). By contrast, lower-status group members demonstrated outright outgroup preference in several domains, indicated by negative average D-scores. Where they did show ingroup preference, it was substantially attenuated compared to higher-status groups. Welch's *t*-tests confirmed that higher-status group members produced significantly higher D-scores than lower-status group members in all six domains (all *p*s < .001; Table S3). The magnitude of this difference in ingroup preference between higher-status and lower-status groups was largest for the weight ($d = 2.39$), age ($d = 2.25$), disability ($d = 2.15$), and skin tone ($d = 1.08$) domains. The gap was smaller, though still reliable, for race ($d = 0.73$) and sexuality ($d = 0.33$), domains where lower-status groups maintained an attenuated ingroup preference. Substantively, this translates to: For sexuality and race, higher-status respondents (straight, White) showed strong ingroup-affirming behavior, while lower-status respondents (gay/lesbian, Black) showed weak ingroup-affirming behavior;

but for weight, disability, age, and skin tone domains, both higher-status and lower-status respondents demonstrated a preference for the higher-status group, albeit with lower-status respondents typically demonstrating a weaker preference for the higher-status group than their higher-status counterparts.

A note of caution is warranted when comparing D-score patterns to QPM parameter estimates. Although both metrics are derived from the same IAT data, they draw on different aspects of task performance: the D-score is computed from response latencies only, whereas QPM parameters are estimated from the frequencies of correct and incorrect responses across eight unique trial types, with no latency information used to estimate the parameters. This means the two metrics are not directly comparable — a given D-score pattern is consistent with multiple distinct configurations of QPM parameters, and agreement or disagreement between them should not be interpreted as direct validation or contradiction. The D-score analyses presented here are intended to characterize the overall behavioral pattern, alongside the model comparison analyses, that the QPM parameters seek to explain; the D-score analyses are not meant to serve as an independent test of the process-level conclusions.

Table S2

Descriptive Statistics for IAT D-Scores by Domain and Group Status

	Higher-Status			Lower-Status		
	N	*M*	*SD*	N	*M*	*SD*
Sexuality	800,088	0.34	0.45	107,272	0.20	0.44
Race	2,400,743	0.35	0.42	375,023	0.05	0.44
Skin Tone	526,679	0.36	0.40	130,283	-0.08	0.45
Age	282,488	0.42	0.38	195,675	-0.46	0.40
Disability	283,233	0.53	0.44	60,099	-0.44	0.48
Weight	278,748	0.52	0.39	223,628	-0.44	0.42

Note. To allow for direct comparison, lower-status respondents' D-scores were reverse-coded prior to analysis. Consequently, for all respondents, positive values indicate faster responding consistent with ingroup preference. *M* = mean; *SD* = standard deviation. D-scores could not be computed for 734 sessions (Race = 250, Skin Tone = 389, Age = 95), which are excluded here; all other analyses use the full analytic sample.

Table S3

T-Tests Comparing IAT D-Scores Between Higher-Status and Lower-Status Group Members

	ΔM	*t*	*df*	*p*	*d*	95% CI
Sexuality	0.15	100.99	138,452	< .001	0.33	[0.32, 0.33]
Race	0.31	403.58	488,527	< .001	0.73	[0.73, 0.74]
Skin Tone	0.44	325.73	184,643	< .001	1.08	[1.07, 1.09]
Age	0.88	756.31	403,667	< .001	2.25	[2.24, 2.26]
Disability	0.96	451.01	82,745	< .001	2.15	[2.14, 2.16]
Weight	0.96	833.92	463,330	< .001	2.39	[2.38, 2.39]

Note. ΔM = mean D-score difference (higher-status – lower-status). Positive values indicate higher D-scores for higher-status group members. *df* = Welch–Satterthwaite degrees of freedom. Cohen's *d* computed using pooled SD; 95% CI based on the normal approximation to the sampling distribution of *d*.

The Quadruple Process Model

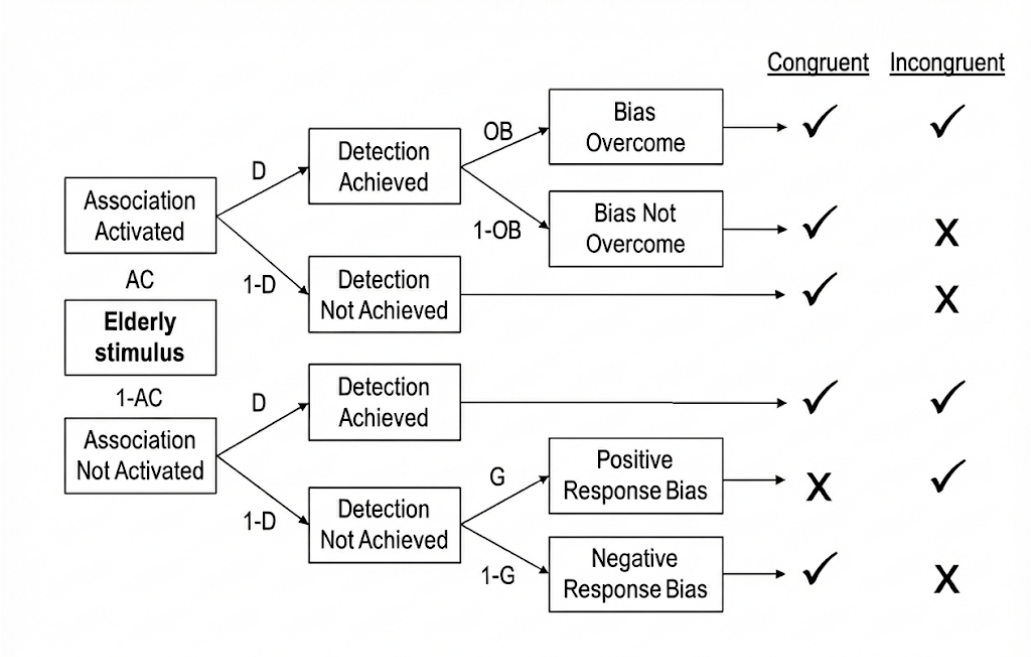
The Quadruple Process Model (QPM) is a multinomial processing tree (MPT) model that quantifies task performance via the joint contribution of four latent cognitive processes: Association Activation (AC), Detection (D), Overcoming Bias (OB), and Guessing (G). As illustrated in Figure S1, the model assumes that a response on any single trial is the outcome of one of several possible cognitive paths. The total probability of a correct response (checkmark) is the sum of the probabilities of all paths leading to it. The total probability of an incorrect response (X) is the sum of the probabilities of all paths leading to it. These probabilities differ depending on whether the trial is Congruent or Incongruent. Congruent trials are those where the evaluative association (e.g., *Elderly = Bad*) facilitates the correct response (e.g., pressing the "Bad" key whenever an image of an elderly person appears). In this case, even if the participant fails to overcome the bias (1-OB), the response will still be correct because the bias aligns with the task requirement. That is, OB essentially drops out of the equation for congruent trials because the response is correct regardless of whether self-regulation succeeds or fails.

$$P(\text{Correct} \mid \text{Congruent}) = [AC \times D] + [(1 - AC) \times D] + [(1 - AC) \times (1 - D) \times G]$$

Incongruent trials are those where the evaluative association conflicts with the correct response (e.g., pressing the "Good" key whenever an image of an elderly person appears, if the association is *Elderly = Bad*). Here, a correct response requires either that the association was not activated or that the association was activated but successfully overcome (OB).

$$P(\text{Correct} \mid \text{Incongruent}) = [(1 - AC) \times D] + [AC \times D \times OB] + [(1 - AC) \times (1 - D) \times G]$$

Figure S1



Note. The processing tree illustrates the probabilistic cognitive paths that lead to a response in the Quadruple Process Model. Each path represents a sequence of cognitive states. For example, the top branch represents a trial where an association is activated (AC), the correct response is detected (D), and the bias is successfully overcome (OB). The columns on the right indicate whether a given path results in a correct (checkmark) or incorrect (X) response, depending on whether the trial is *Congruent* (the association matches the correct response) or *Incongruent* (the association conflicts with the correct response).

Ingroup-Affirming vs. Outgroup-Affirming Specifications

To capture the direction of respondents' intergroup bias (i.e., their evaluative associations), we specified the QPM with two variants. While the mathematical equations above remain identical, the definition of which trials are "Congruent" versus "Incongruent" and, consequently, what the parameters represent, changes based on the hypothesized direction of the bias.

The *ingroup-affirming specification* assumes the participant holds a bias favoring their ingroup and disfavoring the outgroup. Under this specification, congruent trials are defined as IAT blocks where the ingroup shares a key with "Good" and the outgroup shares a key with

"Bad" (e.g., for a Young participant, *Young+Good* / *Elderly+Bad*). Conversely, incongruent trials are those where the mapping is reversed (e.g., for a Young participant, *Elderly+Good* / *Young+Bad*). Consequently, the AC parameter encodes the frequency of activating positive ingroup associations and negative outgroup associations, and the OB parameter encodes the ability to overcome these specific associations when the task requires it.

The *outgroup-affirming specification* assumes the participant holds a reversed bias, favoring the outgroup and disfavoring the ingroup. Under this specification, congruent trials are defined as IAT blocks where the outgroup shares a key with "Good" and the ingroup shares a key with "Bad" (e.g., for a young participant, *Elderly+Good / Young+Bad*). Incongruent trials are defined as blocks where the mapping is reversed (e.g., *Young+Good / Elderly+Bad*). Accordingly, the AC parameter encodes the frequency of activating negative ingroup associations and positive outgroup associations, and the OB parameter encodes the ability to overcome these specific associations when the task requires it.

Model Specification and Comparison

To determine the predominant direction of intergroup bias for each respondent, we compared the fit of the two QPM specifications to each respondent's data, excluding those whose datasets fit identically between the two models (i.e., ties). The tables below reflect model fits post-exclusion. Table S4 summarizes the prevalence of these model fits within each group. The *p-value* column reports the results of exact binomial tests comparing the proportion of outgroup-affirming fits against a chance baseline of 50%. A significant result indicates that the group's preference is systematically skewed toward either the ingroup or outgroup. The *Odds* column quantifies this skew as the ratio of outgroup-affirming to ingroup-affirming fits; values greater than 1.0 indicate a predominant outgroup preference (i.e., more than 50% of the sample was best fit by the outgroup-affirming model), while values less than 1.0 indicate a predominant ingroup preference.

Table S5 formally tests whether the likelihood of demonstrating outgroup-affirming bias differs significantly between lower-status and higher-status groups. We employed Chi-square tests of independence to compare the distributions of model fits across status groups. A significant value indicates that the rate of outgroup-affirming bias is dependent on group status. To interpret the magnitude and direction of this difference, we report the relative odds that a lower-status respondent will fit the outgroup-affirming model compared to a higher-status respondent. ORs greater than 1.0 indicates that lower-status group members are significantly more likely to demonstrate outgroup-affirming bias than their higher-status counterparts.

Table S4

Proportion of Outgroup-Affirming Model Fits by IAT and Group Status, and Binomial Tests

IAT	Group Status	*N*	$n_{outgroup\text{-}affirming}$	$\%_{outgroup\text{-}affirming}$	Odds	*p*
Sexuality	Higher-Status	800,088	228,863	28.60	0.40	< .001
	Lower-Status	107,272	43,436	40.49	0.68	< .001
Race	Higher-Status	2,400,946	656,485	27.34	0.38	< .001
	Lower-Status	375,070	168,921	45.04	0.82	< .001
Skin Tone	Higher-Status	527,015	139,339	26.44	0.36	< .001
	Lower-Status	130,336	67,792	52.01	1.08	< .001
Age	Higher-Status	282,532	76,837	27.20	0.37	< .001
	Lower-Status	195,726	148,151	75.69	3.11	< .001
Disability	Higher-Status	283,233	74,523	26.31	0.36	< .001
	Lower-Status	60,099	41,874	69.68	2.30	< .001
Weight	Higher-Status	278,748	69,284	24.86	0.33	< .001
	Lower-Status	223,628	157,825	70.57	2.40	< .001

Note. Odds reflect the ratio of outgroup-affirming to ingroup-affirming model fits within each group. All *p*-values are from exact binomial tests comparing the proportion of outgroup-affirming fits to chance (50%).

Table S5

Chi-Square Tests of Independence: Group Status Differences in Outgroup-Affirming Model Fits

IAT	χ^2	*df*	OR	95% CI	*p*
Sexuality	6363.08	1	1.70	[1.68, 1.72]	< .001
Race	48612.92	1	2.18	[2.16, 2.19]	< .001
Skin Tone	31667.31	1	3.02	[2.98, 3.05]	< .001
Age	109161.30	1	8.34	[8.23, 8.45]	< .001
Disability	41603.55	1	6.43	[6.31, 6.56]	< .001
Weight	104707.74	1	7.25	[7.16, 7.34]	< .001

Note. Chi-square tests of independence compared the rates of outgroup-affirming versus ingroup-affirming model fits between higher-status and lower-status groups within each IAT domain. OR = odds ratio reflecting the relative odds of outgroup-affirming fits for lower-status compared to higher-status groups. CI = confidence interval.

Parameter Analyses

We utilized likelihood ratio tests to formally test for differences between parameters. This approach involves comparing the goodness-of-fit of a baseline "unconstrained" model—where parameters are estimated freely—against a nested "constrained" model where the relevant parameters are forced to be equal. The difference in model fit is calculated as ΔG^2. A statistically significant ΔG^2 indicates that imposing the equality constraint significantly reduced the model's ability to fit the data, thereby demonstrating that the parameters are significantly different. To assess the magnitude of these differences independent of the large sample size, we also report the change in Cohen's *w* effect size (Δw) and the ratio of the constrained to unconstrained *w*. The latter metric is recommended for comparison effect sizes. This metric quantifies the *relative* degradation in model fit caused by the constraint (e.g., a ratio of 1.5 indicates that the constraint increased misfit by 50%), providing a clearer index of the parameter difference than the absolute Δw alone

Table S6

Equality Constraints on Ingroup Association Activation (AC)

IAT	Group Status	ΔG^2	*df*	*p*	Δw	*w* Ratio
Sexuality	Higher-Status	87856.6	1	< .001	.0092	1.22
	Lower-Status	26936.4	1	< .001	.0178	1.37
Race	Higher-Status	44538.0	1	< .001	.0035	1.17
	Lower-Status	10887.7	1	< .001	.0060	1.33
Skin Tone	Higher-Status	18616.2	1	< .001	.0058	1.27
	Lower-Status	216.9	1	< .001	.0005	1.04
Age	Higher-Status	15466.7	1	< .001	.0129	2.20
	Lower-Status	6770.5	1	< .001	.0093	1.94
Disability	Higher-Status	21311.4	1	< .001	.0082	1.24
	Lower-Status	1383.5	1	< .001	.0027	1.08
Weight	Higher-Status	17543.3	1	< .001	.0109	1.62
	Lower-Status	5446.6	1	< .001	.0043	1.20

Note. Results of likelihood ratio tests comparing models with AC parameters for ingroup targets freely estimated across ingroup-affirming and outgroup-affirming specifications versus models with AC constrained to equality. ΔG^2 = change in likelihood ratio chi-square; Δw = change in Cohen's w effect size; w Ratio = ratio of constrained to unconstrained w (values > 1 indicate worse fit when parameters are constrained to equality).

Table S7

Equality Constraints on Outgroup Association Activation (AC)

IAT	Group Status	ΔG^2	*df*	*p*	Δw	*w* Ratio
Sexuality	Higher-Status	3791.5	1	< .001	.0005	1.01
	Lower-Status	253.9	1	< .001	< .0001	1.00
Race	Higher-Status	65701.6	1	< .001	.0049	1.25
	Lower-Status	11629.5	1	< .001	.0062	1.35
Skin Tone	Higher-Status	17822.7	1	< .001	.0058	1.27
	Lower-Status	560.2	1	< .001	.0013	1.10
Age	Higher-Status	13311.4	1	< .001	.0115	2.07
	Lower-Status	9208.8	1	< .001	.0119	2.20
Disability	Higher-Status	12402.0	1	< .001	.0051	1.15
	Lower-Status	3474.1	1	< .001	.0064	1.18
Weight	Higher-Status	16569.3	1	< .001	.0105	1.60
	Lower-Status	5483.2	1	< .001	.0043	1.20

Note. Results of likelihood ratio tests comparing models with AC parameters for outgroup targets freely estimated across ingroup-affirming and outgroup-affirming specifications versus models with AC constrained to equality. ΔG^2 = change in likelihood ratio chi-square; Δw = change in Cohen's w effect size; w Ratio = ratio of constrained to unconstrained w (values > 1 indicate worse fit when parameters are constrained to equality).

Table S8

Equality Constraints on the Detection of Correct Responses (D)

IAT	Group Status	ΔG^2	*df*	*p*	Δw	*w* Ratio
Sexuality	Higher-Status	17890.4	1	< .001	.0022	1.05
	Lower-Status	581.3	1	< .001	.0009	1.02
Race	Higher-Status	105078.0	1	< .001	.0079	1.40
	Lower-Status	192.8	1	< .001	< .0001	1.00
Skin Tone	Higher-Status	24118.0	1	< .001	.0084	1.39
	Lower-Status	1010.0	1	< .001	.0021	1.16
Age	Higher-Status	7968.1	1	< .001	.0082	1.76
	Lower-Status	5594.0	1	< .001	.0091	1.91
Disability	Higher-Status	14175.9	1	< .001	.0064	1.19
	Lower-Status	3351.4	1	< .001	.0067	1.18
Weight	Higher-Status	9510.9	1	< .001	.0070	1.40
	Lower-Status	8737.7	1	< .001	.0069	1.33

Note. Results of likelihood ratio tests comparing models with D parameters freely estimated across ingroup-affirming and outgroup-affirming specifications versus models with D constrained to equality. ΔG^2 = change in likelihood ratio chi-square; Δw = change in Cohen's w effect size; w Ratio = ratio of constrained to unconstrained w (values > 1 indicate worse fit when parameters are constrained to equality).

Table S9

Equality Constraints on Guessing (G)

IAT	Group Status	ΔG^2	*df*	*p*	Δw	*w* Ratio
Sexuality	Higher-Status	1495.9	1	< .001	.0003	1.01
	Lower-Status	1498.3	1	< .001	.0017	1.04
Race	Higher-Status	9938.3	1	< .001	.0011	1.05
	Lower-Status	3.0	1	0.081	< .0001	1.00
Skin Tone	Higher-Status	2302.2	1	< .001	.0012	1.05
	Lower-Status	30.3	1	< .001	< .0001	1.01
Age	Higher-Status	70.2	1	< .001	< .0001	1.01
	Lower-Status	476.0	1	< .001	.0010	1.10
Disability	Higher-Status	160.0	1	< .001	.0002	1.01
	Lower-Status	20.4	1	< .001	.0001	1.00
Weight	Higher-Status	86.1	1	< .001	.0001	1.01
	Lower-Status	21.5	1	< .001	< .0001	1.00

Note. Results of likelihood ratio tests comparing models with G parameters freely estimated across ingroup-affirming and outgroup-affirming specifications versus models with G constrained to equality. ΔG^2 = change in likelihood ratio chi-square; Δw = change in Cohen's w effect size; w Ratio = ratio of constrained to unconstrained w (values > 1 indicate worse fit when parameters are constrained to equality).

Table S10

Equality Constraints on Overcoming Bias (OB)

IAT	Group Status	ΔG^2	*df*	*p*	Δw	*w* Ratio
Sexuality	Higher-Status	3050.4	1	< .001	.0004	1.01
	Lower-Status	13369.8	1	< .001	.0103	1.21
Race	Higher-Status	24286.6	1	< .001	.0020	1.10
	Lower-Status	32362.5	1	< .001	.0143	1.79
Skin Tone	Higher-Status	371.3	1	< .001	.0002	1.01
	Lower-Status	7131.1	1	< .001	.0117	1.88
Age	Higher-Status	741.2	1	< .001	.0010	1.09
	Lower-Status	1371.3	1	< .001	.0026	1.26
Disability	Higher-Status	0.0	1	1.000	< .0001	1.00
	Lower-Status	1.3	1	0.247	< .0001	1.00
Weight	Higher-Status	17.9	1	< .001	< .0001	1.00
	Lower-Status	4147.1	1	< .001	.0033	1.16

Note. Results of likelihood ratio tests comparing models with OB parameters freely estimated across ingroup-affirming and outgroup-affirming specifications versus models with OB constrained to equality. ΔG^2 = change in likelihood ratio chi-square; Δw = change in Cohen's w effect size; w Ratio = ratio of constrained to unconstrained w (values > 1 indicate worse fit when parameters are constrained to equality).

Parameter Estimates

To aid in the interpretation of the following tables, it is useful to benchmark the typical magnitude and scaling of QPM parameters. The Detection (D) and Association Activation (AC) parameters represent the probabilities of these respective cognitive events occurring, scaled from 0 (absence) to 1 (certainty). In standard IAT data, D estimates are typically high (often > .80), reflecting that participants generally identify stimuli and task rules correctly. If D = .89, it suggests that respondents identify the correct response 89% of the time. AC estimates are typically lower (often < .15), indicating that evaluative associations are not activated on most trials. An estimate of ingroup-positive AC = .08 suggests that respondents activated a positive association of their ingroup on 8% of trials in which an image of their ingroup was presented.

In contrast, the Response Bias (G) and Overcoming Bias (OB) parameters are best interpreted relative to a benchmark of .50 rather than zero. For G, a value of .50 indicates neutral, or no response bias between the two response keys (i.e., no bias toward the "good" key or "bad" key), while deviations imply a tendency toward a specific key. For OB, a value of .50 indicates that self-regulation succeeds as often as it fails. Values substantially above .50 indicate effective control, whereas values below .50 indicate that evaluative associations typically win out over the detected correct response.

Table S11

Association Activation (AC) Parameter Estimates for Higher-Status Groups

IAT	Model	*AC* (Ingroup)	95% CI	*AC* (Outgroup)	95% CI
Sexuality	Ingroup-Affirming	0.095	[0.095, 0.095]	0.050	[0.050, 0.050]
	Outgroup-Affirming	0.036	[0.036, 0.037]	0.039	[0.039, 0.039]
Race	Ingroup-Affirming	0.086	[0.086, 0.086]	0.056	[0.056, 0.056]
	Outgroup-Affirming	0.059	[0.059, 0.060]	0.028	[0.028, 0.028]
Skin Tone	Ingroup-Affirming	0.101	[0.100, 0.101]	0.071	[0.070, 0.071]
	Outgroup-Affirming	0.061	[0.061, 0.062]	0.035	[0.035, 0.036]
Age	Ingroup-Affirming	0.095	[0.095, 0.096]	0.071	[0.071, 0.072]
	Outgroup-Affirming	0.045	[0.045, 0.046]	0.030	[0.030, 0.031]
Disability	Ingroup-Affirming	0.092	[0.092, 0.093]	0.052	[0.052, 0.052]
	Outgroup-Affirming	0.044	[0.043, 0.044]	0.021	[0.021, 0.022]
Weight	Ingroup-Affirming	0.099	[0.099, 0.099]	0.075	[0.074, 0.075]
	Outgroup-Affirming	0.047	[0.046, 0.047]	0.028	[0.028, 0.029]

Note. AC = Association Activation parameter estimates from the Quadruple Process Model. For ingroup-affirming models, AC (Ingroup) reflects positive associations with the ingroup and AC (Outgroup) reflects negative associations with the outgroup. For outgroup-affirming models, AC (Ingroup) reflects negative associations with the ingroup and AC (Outgroup) reflects positive associations with the outgroup. CI = confidence interval.

Table S12

Association Activation (AC) Parameter Estimates for Lower-Status Groups

IAT	Model	*AC* (Ingroup)	95% CI	*AC* (Outgroup)	95% CI
Sexuality	Ingroup-Affirming	0.089	[0.089, 0.090]	0.076	[0.075, 0.077]
	Outgroup-Affirming	0.014	[0.014, 0.015]	0.068	[0.067, 0.069]
Race	Ingroup-Affirming	0.072	[0.072, 0.072]	0.084	[0.084, 0.084]
	Outgroup-Affirming	0.045	[0.044, 0.045]	0.054	[0.054, 0.055]
Skin Tone	Ingroup-Affirming	0.069	[0.069, 0.070]	0.076	[0.075, 0.077]
	Outgroup-Affirming	0.062	[0.062, 0.063]	0.064	[0.064, 0.065]
Age	Ingroup-Affirming	0.027	[0.026, 0.027]	0.032	[0.031, 0.032]
	Outgroup-Affirming	0.055	[0.054, 0.055]	0.069	[0.068, 0.069]
Disability	Ingroup-Affirming	0.031	[0.030, 0.032]	0.054	[0.052, 0.055]
	Outgroup-Affirming	0.056	[0.055, 0.056]	0.099	[0.098, 0.100]
Weight	Ingroup-Affirming	0.033	[0.032, 0.033]	0.050	[0.050, 0.051]
	Outgroup-Affirming	0.059	[0.059, 0.060]	0.079	[0.079, 0.080]

Note. AC = Association Activation parameter estimates from the Quadruple Process Model. For ingroup-affirming models, AC (Ingroup) reflects positive associations with the ingroup and AC (Outgroup) reflects negative associations with the outgroup. For outgroup-affirming models, AC (Ingroup) reflects negative associations with the ingroup and AC (Outgroup) reflects positive associations with the outgroup. CI = confidence interval.

Table S13

Detection (D) Parameter Estimates for Higher-Status Groups

IAT	Model	*D*	95% CI
Sexuality	Ingroup-Affirming	0.867	[0.867, 0.867]
	Outgroup-Affirming	0.846	[0.846, 0.846]
Race	Ingroup-Affirming	0.921	[0.921, 0.922]
	Outgroup-Affirming	0.897	[0.897, 0.897]
Skin Tone	Ingroup-Affirming	0.915	[0.915, 0.915]
	Outgroup-Affirming	0.888	[0.888, 0.889]
Age	Ingroup-Affirming	0.863	[0.863, 0.863]
	Outgroup-Affirming	0.838	[0.838, 0.839]
Disability	Ingroup-Affirming	0.899	[0.898, 0.899]
	Outgroup-Affirming	0.869	[0.869, 0.870]
Weight	Ingroup-Affirming	0.903	[0.903, 0.903]
	Outgroup-Affirming	0.879	[0.878, 0.879]

Note. D = Detection parameter estimates from the Quadruple Process Model, reflecting the probability of correctly identifying the stimulus and task requirements. CI = confidence interval.

Table S14

Detection (D) Parameter Estimates for Lower-Status Groups

IAT	Model	*D*	95% CI
Sexuality	Ingroup-Affirming	0.864	[0.864, 0.865]
	Outgroup-Affirming	0.873	[0.873, 0.874]
Race	Ingroup-Affirming	0.896	[0.896, 0.896]
	Outgroup-Affirming	0.899	[0.898, 0.899]
Skin Tone	Ingroup-Affirming	0.881	[0.881, 0.882]
	Outgroup-Affirming	0.892	[0.891, 0.892]
Age	Ingroup-Affirming	0.938	[0.938, 0.939]
	Outgroup-Affirming	0.954	[0.954, 0.955]
Disability	Ingroup-Affirming	0.861	[0.860, 0.861]
	Outgroup-Affirming	0.891	[0.891, 0.892]
Weight	Ingroup-Affirming	0.894	[0.893, 0.894]
	Outgroup-Affirming	0.917	[0.916, 0.917]

Note. D = Detection parameter estimates from the Quadruple Process Model, reflecting the probability of correctly identifying the stimulus and task requirements. CI = confidence interval.

Table S15

Guessing (G) Parameter Estimates for Higher-Status Groups

IAT	Model	*G*	95% CI
Sexuality	Ingroup-Affirming	0.533	[0.532, 0.533]
	Outgroup-Affirming	0.551	[0.550, 0.551]
Race	Ingroup-Affirming	0.557	[0.556, 0.557]
	Outgroup-Affirming	0.522	[0.521, 0.523]
Skin Tone	Ingroup-Affirming	0.555	[0.554, 0.555]
	Outgroup-Affirming	0.520	[0.519, 0.521]
Age	Ingroup-Affirming	0.528	[0.527, 0.529]
	Outgroup-Affirming	0.521	[0.520, 0.523]
Disability	Ingroup-Affirming	0.535	[0.534, 0.536]
	Outgroup-Affirming	0.546	[0.545, 0.548]
Weight	Ingroup-Affirming	0.522	[0.521, 0.523]
	Outgroup-Affirming	0.513	[0.511, 0.514]

Note. G = Guessing parameter estimates from the Quadruple Process Model, reflecting response bias when no association is activated and the correct answer is not detected. CI = confidence interval.

Table S16

Guessing (G) Parameter Estimates for Lower-Status Groups

IAT	Model	*G*	95% CI
Sexuality	Ingroup-Affirming	0.594	[0.592, 0.595]
	Outgroup-Affirming	0.546	[0.544, 0.548]
Race	Ingroup-Affirming	0.552	[0.551, 0.553]
	Outgroup-Affirming	0.551	[0.550, 0.552]
Skin Tone	Ingroup-Affirming	0.539	[0.537, 0.541]
	Outgroup-Affirming	0.546	[0.544, 0.547]
Age	Ingroup-Affirming	0.568	[0.565, 0.570]
	Outgroup-Affirming	0.604	[0.602, 0.606]
Disability	Ingroup-Affirming	0.548	[0.545, 0.551]
	Outgroup-Affirming	0.540	[0.538, 0.542]
Weight	Ingroup-Affirming	0.519	[0.518, 0.521]
	Outgroup-Affirming	0.524	[0.523, 0.526]

Note. G = Guessing parameter estimates from the Quadruple Process Model, reflecting response bias when no association is activated and the correct answer is not detected. CI = confidence interval.

Table S17

Overcoming Bias (OB) Parameter Estimates for Higher-Status Groups

IAT	Model	*OB*	95% CI
Sexuality	Ingroup-Affirming	0.115	[0.112, 0.119]
	Outgroup-Affirming	0.000	[-0.009, 0.009]
Race	Ingroup-Affirming	0.434	[0.433, 0.435]
	Outgroup-Affirming	0.725	[0.722, 0.728]
Skin Tone	Ingroup-Affirming	0.771	[0.768, 0.773]
	Outgroup-Affirming	0.837	[0.831, 0.843]
Age	Ingroup-Affirming	0.363	[0.359, 0.368]
	Outgroup-Affirming	0.133	[0.116, 0.151]
Disability	Ingroup-Affirming	0.000	[-0.005, 0.005]
	Outgroup-Affirming	0.000	[-0.017, 0.017]
Weight	Ingroup-Affirming	0.479	[0.475, 0.483]
	Outgroup-Affirming	0.509	[0.496, 0.522]

Note. OB = Overcoming Bias parameter estimates from the Quadruple Process Model, reflecting the probability of overcoming an activated association to respond correctly. For ingroup-affirming models, OB reflects the overcoming of positive ingroup and negative outgroup associations. For outgroup-affirming models, OB reflects the overcoming of negative ingroup and positive outgroup associations. CI = confidence interval.

Table S18

Overcoming Bias (OB) Parameter Estimates for Lower-Status Groups

IAT	Model	*OB*	95% CI
Sexuality	Ingroup-Affirming	0.675	[0.668, 0.682]
	Outgroup-Affirming	0.000	[-0.016, 0.016]
Race	Ingroup-Affirming	0.788	[0.785, 0.791]
	Outgroup-Affirming	0.122	[0.114, 0.130]
Skin Tone	Ingroup-Affirming	0.825	[0.818, 0.831]
	Outgroup-Affirming	0.350	[0.341, 0.359]
Age	Ingroup-Affirming	0.616	[0.601, 0.630]
	Outgroup-Affirming	0.296	[0.291, 0.301]
Disability	Ingroup-Affirming	0.019	[-0.013, 0.050]
	Outgroup-Affirming	0.000	[-0.010, 0.010]
Weight	Ingroup-Affirming	0.633	[0.622, 0.644]
	Outgroup-Affirming	0.175	[0.169, 0.180]

Note. OB = Overcoming Bias parameter estimates from the Quadruple Process Model, reflecting the probability of overcoming an activated association to respond correctly. For ingroup-affirming models, OB reflects the overcoming of positive ingroup and negative outgroup associations. For outgroup-affirming models, OB reflects the overcoming of negative ingroup and positive outgroup associations. CI = confidence interval.

Results of the Detection (D) and Guessing (G) Parameters

Formal testing of the D and G parameters across conditions relied on the same equality constraints approach used to examine the AC and OB parameters. That is, likelihood-ratio tests compared whether model fit significantly worsened as a function constraining D or G to be equal across ingroup-affirming and outgroup-affirming model specifications, for both lower-status and higher-status groups. We concluded that parameters differed significantly if the equality-constrained model produced a significantly worse goodness-of-fit (G2) compared to the unconstrained model.

Detection (D)

The analyses revealed that the unconstrained model provided significantly better fit for the Detection (D) parameter in all domains (Table S8). A consistent directional pattern characterized the D differences: D was higher under whichever specification aligned with culturally dominant attitudes. For higher-status groups, D was higher under the ingroup-affirming specification across all six domains (Table S13). For lower-status groups, D was higher under the outgroup-affirming specification across all six domains (Table S14).

Research on D indicates that it reflects a controlled, capacity-dependent process for detecting correct responses: response deadlines selectively reduce D (2, 3), and neuroimaging data associate D with anterior cingulate, dorsolateral prefrontal cortex, and operculum activity during conditions requiring discrimination between competing responses (4). D is also sensitive to the content of practiced stimulus-response pairings rather than practice alone (5). Furthermore, deliberate regulatory strategies (implementation intentions) increase OB but decrease D (6), a double dissociation indicating that D reflects control-oriented processing that is functionally distinct from the self-regulatory control captured by OB. Unlike OB, which showed a status-

specific asymmetry (i.e., higher ingroup-affirming versus outgroup-affirming OB, among lower-status groups), the D pattern was symmetric across group status: both higher-status and lower-status groups showed elevated D under the culturally dominant specification. The D pattern reported here is therefore unlikely to reflect the same self-regulatory process that is the focus of the present paper.

Why it is that D is consistently higher under the culturally dominant specification remains an open question. D is responsive to training with specific stimulus-response pairings (5), raising the possibility that those whose evaluative associations align with culturally dominant attitudes may also navigate their social environments in ways that more frequently demand discrimination of domain-relevant stimuli in the corresponding direction. Because both higher- and lower-status groups are embedded in the same stratified cultural environment, both would be expected to show elevated D under the culturally dominant specification, as observed in the current data. However, the experimental evidence is not straightforwardly consistent with this account: Calanchini and colleagues (5) found that D was selectively enhanced by counter-prejudicial training rather than pro-prejudicial training. Future research should directly test the mechanisms linking evaluative learning histories to variation in detection across model specifications.

Guessing (G)

The Guessing (G) parameter showed statistically significant constraint violations in most domains (Table S9), but no consistent directional pattern emerged across group status, model specification, or attitude domain. These results may reflect the high sensitivity of the likelihood-ratio test at the present sample sizes rather than psychologically meaningful variation in guessing behavior.

Model Fit

Model fit statistics were used to evaluate how well the Quadruple Process Model accounts for the observed response patterns in the data. In multinomial processing tree models like the Quad Model, the likelihood ratio statistic (G^2) tests whether there is a significant discrepancy between the response frequencies predicted by the model and those actually observed. Therefore, unlike most statistical tests where a significant result is desirable, a non-significant G^2 is preferred here because it indicates that the model's predictions closely match the observed data.

The G^2 statistic asymptotically follows a chi-square distribution, which makes it highly sensitive to sample size. As sample size increases, even trivially small deviations in model misfit become statistically detectable. Given the very large sample sizes analyzed in the present study (millions of IAT trials per domain), statistically significant G^2 values are expected even when deviations from the model are negligible in practical terms. For this reason, we also report Cohen's w, an effect size measure that quantifies the magnitude of misfit, adjusting for sample size. Cohen's w is calculated as $w = \sqrt{(G^2/N)}$, where N is the number of observations. Following conventional guidelines, $w < .10$ indicates small misfit, $w = .30$ indicates medium misfit, and $w = .50$ indicates large misfit. As shown in Table S19, although the G^2 statistics are statistically significant for all models, the w values fall well below the .10 threshold (range: .010 – .048), indicating excellent practical fit.

Table S19

Model Fit Statistics by IAT and Group Status

IAT	Group Status	G^2	*df*	*p*	*w*
Sexuality	Higher-Status	172521.7	6	< .001	.0424
	Lower-Status	29530.0	6	< .001	.0481
Race	Higher-Status	115832.5	6	< .001	.0199
	Lower-Status	14758.9	6	< .001	.0180
Skin Tone	Higher-Status	29593.0	6	< .001	.0216
	Lower-Status	2756.2	6	< .001	.0133
Age	Higher-Status	3956.1	6	< .001	.0108
	Lower-Status	2310.0	6	< .001	.0099
Disability	Higher-Status	38472.6	6	< .001	.0338
	Lower-Status	9412.5	6	< .001	.0363
Weight	Higher-Status	10404.1	6	< .001	.0175
	Lower-Status	12114.6	6	< .001	.0211

Note. Model fit statistics for the Quadruple Process Model with separate specifications for ingroup-affirming and outgroup-affirming response patterns. G^2 = likelihood ratio chi-square statistic; w = Cohen's w effect size measure of practical fit.

Quadruple Process Model with Two OB Parameters

The standard QPM assumes a single OB parameter, implying that the capacity to overcome activated associations is uniform across association types (e.g., ingroup-affirming associations: positive associations of the ingroup, negative associations of the outgroup). To test whether self-regulatory control differed across specific types of associations, we refit the QPM with a separate OB for each association (2). This specification allows for a more granular analysis, determining, for instance, whether the self-regulatory control of ingroup-affirming evaluations is driven primarily by control over positive ingroup evaluations or negative outgroup evaluations.

Table S20

Within-Model Equality Constraints on Overcoming Bias (OB) for Higher-Status Groups

IAT	Model	ΔG^2	*df*	*p*	Δw	*w* Ratio
Sexuality	Ingroup-Affirming	4100.4	1	< .001	.0005	1.01
	Outgroup-Affirming	0.0	1	1.000	< .0001	1.00
Race	Ingroup-Affirming	956.3	1	< .001	< .0001	1.00
	Outgroup-Affirming	327.9	1	< .001	< .0001	1.00
Skin Tone	Ingroup-Affirming	1586.1	1	< .001	.0005	1.02
	Outgroup-Affirming	524.3	1	< .001	.0003	1.01
Age	Ingroup-Affirming	51.8	1	< .001	< .0001	1.01
	Outgroup-Affirming	140.1	1	< .001	.0002	1.02
Disability	Ingroup-Affirming	0.0	1	1.000	< .0001	1.00
	Outgroup-Affirming	0.0	1	1.000	< .0001	1.00
Weight	Ingroup-Affirming	1384.8	1	< .001	.0011	1.07
	Outgroup-Affirming	171.6	1	< .001	.0001	1.01

Note. Results of likelihood ratio tests comparing models with separate OB parameters for ingroup targets and outgroup targets versus models with these parameters constrained to equality within each model specification. ΔG^2 = change in likelihood ratio chi-square; Δw = change in Cohen's w effect size; w Ratio = ratio of constrained to unconstrained w (values > 1 indicate worse fit when parameters are constrained to equality).

Table S21

Within-Model Equality Constraints on Overcoming Bias (OB) for Lower-Status Groups

IAT	Model	ΔG^2	*df*	*p*	Δw	*w* Ratio
Sexuality	Ingroup-Affirming	1187.7	1	< .001	.0016	1.03
	Outgroup-Affirming	-0.0	1	1.000	< .0001	1.00
Race	Ingroup-Affirming	295.0	1	< .001	.0002	1.01
	Outgroup-Affirming	809.1	1	< .001	.0005	1.03
Skin Tone	Ingroup-Affirming	223.1	1	< .001	.0006	1.04
	Outgroup-Affirming	0.0	1	.989	< .0001	1.00
Age	Ingroup-Affirming	763.6	1	< .001	.0018	1.22
	Outgroup-Affirming	0.8	1	.367	< .0001	1.00
Disability	Ingroup-Affirming	100.8	1	< .001	.0003	1.01
	Outgroup-Affirming	0.0	1	1.000	< .0001	1.00
Weight	Ingroup-Affirming	956.5	1	< .001	.0010	1.05
	Outgroup-Affirming	1360.5	1	< .001	.0011	1.06

Note. Results of likelihood ratio tests comparing models with separate OB parameters for ingroup targets and outgroup targets versus models with these parameters constrained to equality within each model specification. ΔG^2 = change in likelihood ratio chi-square; Δw = change in Cohen's w effect size; w Ratio = ratio of constrained to unconstrained w (values > 1 indicate worse fit when parameters are constrained to equality).

Table S22

Overcoming Bias (OB) Parameter Estimates for Higher-Status Groups

IAT	Model	*OB* (Ingroup)	95% CI	*OB* (Outgroup)	95% CI
Sexuality	Ingroup-Affirming	.034	[.029, .038]	.304	[.298, .310]
	Outgroup-Affirming	.000	[-0.013, .013]	.000	[-0.014, .014]
Race	Ingroup-Affirming	.416	[.414, .417]	.466	[.464, .469]
	Outgroup-Affirming	.742	[.738, .745]	.660	[.652, .668]
Skin Tone	Ingroup-Affirming	.735	[.732, .737]	.827	[.823, .830]
	Outgroup-Affirming	.883	[.876, .890]	.699	[.685, .713]
Age	Ingroup-Affirming	.378	[.372, .384]	.340	[.332, .348]
	Outgroup-Affirming	.218	[.197, .239]	.000	[-0.036, .036]
Disability	Ingroup-Affirming	.000	[-0.006, .006]	.000	[-0.010, .010]
	Outgroup-Affirming	.000	[-0.020, .020]	.000	[-0.042, .042]
Weight	Ingroup-Affirming	.416	[.411, .421]	.567	[.561, .572]
	Outgroup-Affirming	.436	[.418, .454]	.657	[.633, .681]

Note. OB = Overcoming Bias parameter estimates from the two-OB Quadruple Process Model. OB (Ingroup) reflects control over responses to ingroup targets; OB (Outgroup) reflects control over responses to outgroup targets. CI = confidence interval.

Table S23

Overcoming Bias (OB) Parameter Estimates for Lower-Status Groups

IAT	Model	*OB* (Ingroup)	95% CI	*OB* (Outgroup)	95% CI
Sexuality	Ingroup-Affirming	.794	[.785, .803]	.522	[.510, .535]
	Outgroup-Affirming	.000	[-0.065, .065]	.000	[-0.017, .017]
Race	Ingroup-Affirming	.825	[.820, .831]	.762	[.757, .766]
	Outgroup-Affirming	.262	[.251, .273]	.002	[-0.010, .014]
Skin Tone	Ingroup-Affirming	.760	[.749, .771]	.872	[.863, .881]
	Outgroup-Affirming	.350	[.337, .364]	.350	[.337, .364]
Age	Ingroup-Affirming	.299	[.268, .331]	.784	[.768, .800]
	Outgroup-Affirming	.299	[.291, .307]	.294	[.287, .301]
Disability	Ingroup-Affirming	.294	[.242, .345]	.000	[-0.039, .039]
	Outgroup-Affirming	.000	[-0.021, .021]	.000	[-0.012, .012]
Weight	Ingroup-Affirming	.891	[.874, .909]	.483	[.466, .499]
	Outgroup-Affirming	.311	[.303, .320]	.072	[.064, .081]

Note. OB = Overcoming Bias parameter estimates from the two-OB Quadruple Process Model. OB (Ingroup) reflects control over responses to ingroup targets; OB (Outgroup) reflects control over responses to outgroup targets. CI = confidence interval.

Table S24

Two-OB Model Fit Statistics by IAT and Group Status

IAT	Group Status	G^2	*df*	*p*	*w*
Sexuality	Higher-Status	168421.4	4	< .001	.0419
	Lower-Status	28342.3	4	< .001	.0465
Race	Higher-Status	114548.3	4	< .001	.0199
	Lower-Status	13654.8	4	< .001	.0174
Skin Tone	Higher-Status	27482.7	4	< .001	.0208
	Lower-Status	2533.1	4	< .001	.0127
Age	Higher-Status	3764.2	4	< .001	.0105
	Lower-Status	1545.6	4	< .001	.0081
Disability	Higher-Status	38472.6	4	< .001	.0338
	Lower-Status	9311.7	4	< .001	.0360
Weight	Higher-Status	8847.7	4	< .001	.0162
	Lower-Status	9797.7	4	< .001	.0191

Note. Model fit statistics for the two-OB Quadruple Process Model with separate OB parameters for ingroup and outgroup targets. G^2 = likelihood ratio chi-square statistic; w = Cohen's w effect size measure of practical fit.

Individual-Level Parameter Estimation

Methodological Considerations and Data Exclusions

Multinomial processing tree (MPT) models, including the QPM, can be estimated at the aggregate level (as we've reported so far, estimating parameters for data aggregated across subgroups of participants) or at the individual level (estimating parameters separately for each participant). As detailed in the main text, we fit QPMs to each individual participant's data to determine the best-fitting model specification (ingroup-affirming vs. outgroup-affirming). While these individual fits were sufficient for comparing model fit statistics (G^2), we subsequently identified that their parameter estimates were insufficient to serve as the primary basis for analyzing cognitive processes. Individual-level estimation is theoretically appealing because it preserves information about between-person variability. However, individual-level MPT estimation poses substantial challenges when applied to data with low error rates for each individual respondent, as is characteristic of the Implicit Association Test (IAT).

Consider, for instance, that the Overcoming Bias (OB) parameter represents the probability of successfully overcoming an evaluative association in favor of a correct response. By definition, OB is informed by trials in which such a conflict arises; when participants make very few errors (again, as is typical in the IAT), the data provide little information to constrain the OB estimate. This is a structural feature of the model rather than a limitation of the estimation procedure. With insufficient error trials, the likelihood surface along the OB dimension becomes flat and maximum likelihood estimation yields point estimates with extremely wide confidence intervals.

Nevertheless, to evaluate the feasibility of individual-level estimation, we fit the QPM separately to each participant's IAT data and examined the precision of the resulting OB

parameter estimates. We operationalized estimate precision using the width of the 95% confidence interval, calculated as the difference between the upper and lower bounds. Because OB is a probability parameter bounded between 0 and 1, confidence interval widths can range from nearly 0 (almost perfect precision) to 1.0 (complete uncertainty spanning the entire parameter space). We adopted a conservative exclusion criterion, flagging estimates as unreliable if their 95% confidence interval width exceeded .80. That is, if the interval was wider than .80 on the parameter space, with a maximum width of 1.00, we deemed that it provides negligible constraint on the true parameter value and cannot meaningfully distinguish between low, moderate, and high levels of self-regulatory control.

Applying this criterion revealed that the vast majority of individual-level OB estimates failed to meet acceptable precision standards. Across the six IAT domains, approximately 80% of OB parameter estimates exhibited confidence interval widths exceeding the 0.80 threshold. The high exclusion rate reflects the fundamental incompatibility between the low error rates typical of IAT performance and the data requirements for precise OB estimation at the individual level.

Given these constraints, we determined that individual-level OB estimates could not serve as the primary basis for inference in the present research. Retaining only the approximately 20% of estimates meeting our precision criterion (1.1 million OB estimates) would introduce severe selection bias, as participants with sufficient errors to yield precise OB estimates are, by definition, those who performed more poorly on the IAT—a nonrepresentative subset of the population. Moreover, even among retained estimates, the very low precision of the individual-level estimates would limit statistical power for detecting meaningful effects.

We therefore adopted aggregate-level estimation as our primary analytic approach, pooling response frequencies across participants within each combination of IAT domain, group status,

and model specification. Aggregation dramatically increases the effective sample size of errors informing each parameter estimate, yielding highly precise estimates (as evidenced by the narrow confidence intervals reported in the main text and supplementary tables). While this sacrifices information about individual differences, it provides the statistical precision necessary for meaningful inference about the phenomena of central interest: systematic differences in self-regulatory control across group status and model specification. Additionally, large-scale reanalysis of various MPTs have shown that such an aggregated analyses usually results in very similar estimates as fitting more complex, hierarchical models (which is technically infeasible in the present case; 7).

Individual-Level OB Analyses

For transparency, we report the results of individual-level analyses below, acknowledging that these findings should be interpreted with appropriate caution given the substantial data exclusions and the selected nature of the retained sample. Overall, we find a pattern of OB estimates highly similar to that revealed in the aggregate-level parameter analyses.

To test whether the aggregate-level OB pattern replicates at the individual-level, we tested whether individual-level OB estimates varied as a function of participant group status and QPM specification. For each of the six attitude domains, we conducted a 2 (group status: higher vs. lower) × 2 (model type: ingroup-affirming vs. outgroup-affirming) between-subjects analysis of variance (ANOVA) on the OB parameter estimates.

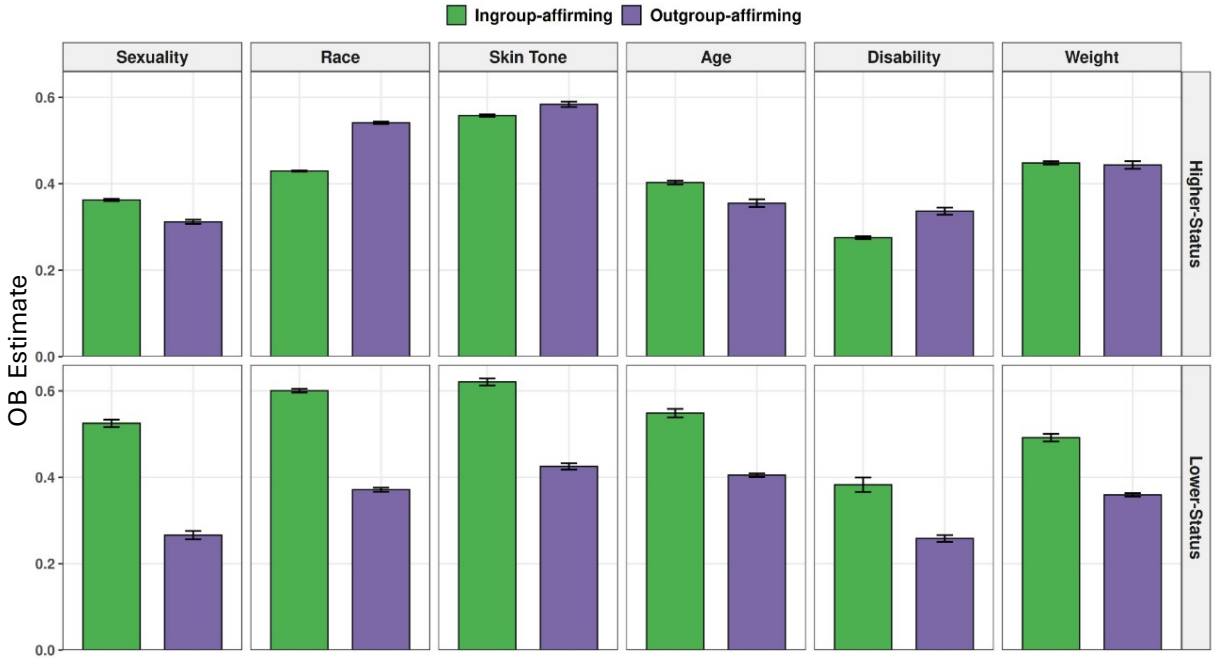
Among lower-status groups, ingroup-affirming evaluations were estimated to be overcome significantly more often than outgroup-affirming evaluations (ps < .001). The effect was strongest for the domains of sexuality (d = 0.62, p < .001), race (d = 0.54, p < .001), and skin

tone ($d = 0.46$, $p < .001$), followed by age ($d = 0.33$, $p < .001$), disability ($d = 0.32$, $p < .001$), and weight ($d = 0.31$, $p < .001$).

Among higher-status groups, the pattern of individual-level OB estimates was inconsistent. In the domains of disability ($d = -0.16$, $p < .001$), race ($d = -0.26$, $p < .001$), and skin tone ($d = -0.06$, $p < .001$), outgroup-affirming evaluations were estimated to be overcome significantly more often than ingroup-affirming evaluations. The opposite pattern was found for the domains of age ($d = 0.11$, $p < .001$) and sexuality ($d = 0.12$, $p < .001$), with a small but reliably greater estimated overcoming of ingroup-affirming evaluations. For the weight IAT, there was no significant difference in the estimated overcoming of ingroup-affirming versus outgroup-affirming evaluations ($d = 0.01$, $p = .325$).

Figure S2

Average Individual-Level OB Estimates by QPM Specification and Group Status



Notes. Error bars represent 95% confidence intervals. Bars reflect the mean of the individual-level Overcoming Bias (OB) parameter estimates. The OB parameter represents the probability of overcoming biased associations in favor of correctly detected responses. Values of OB < .50 indicate that successful self-regulation occurs on less than half of relevant trials, whereas OB > .50 indicates successful self-regulation on more than half of relevant trials. Bar colors denote whether the estimates come from 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).

Table S25

Descriptive Statistics for Individual-Level OB Estimates for Higher-Status Groups

IAT	Model Type	*n*	*M*	*SD*
Sexuality	Ingroup-Affirming	97,982	.363	.411
	Outgroup-Affirming	29,285	.312	.426
Race	Ingroup-Affirming	394,235	.430	.423
	Outgroup-Affirming	102,330	.541	.457
Skin Tone	Ingroup-Affirming	83,149	.558	.416
	Outgroup-Affirming	21,269	.584	.453
Age	Ingroup-Affirming	34,755	.403	.417
	Outgroup-Affirming	9,454	.355	.436
Disability	Ingroup-Affirming	46,119	.275	.383
	Outgroup-Affirming	10,704	.337	.435
Weight	Ingroup-Affirming	45,583	.448	.419
	Outgroup-Affirming	10,359	.444	.460

Note. Descriptive statistics for individual-level Overcoming Bias (OB) parameter estimates. Only estimates meeting the precision criterion (95% CI width < .80) were included. M = mean; SD = standard deviation.

Table S26

Descriptive Statistics for Individual-Level OB Estimates for Lower-Status Groups

IAT	Model Type	*n*	*M*	*SD*
Sexuality	Ingroup-Affirming	10,076	.525	.437
	Outgroup-Affirming	6,576	.266	.395
Race	Ingroup-Affirming	38,932	.600	.425
	Outgroup-Affirming	31,684	.371	.431
Skin Tone	Ingroup-Affirming	10,774	.621	.425
	Outgroup-Affirming	13,006	.425	.433
Age	Ingroup-Affirming	8,827	.549	.468
	Outgroup-Affirming	39,705	.405	.423
Disability	Ingroup-Affirming	2,684	.383	.443
	Outgroup-Affirming	9,096	.259	.376
Weight	Ingroup-Affirming	10,594	.492	.459
	Outgroup-Affirming	37,185	.359	.410

Note. Descriptive statistics for individual-level Overcoming Bias (OB) parameter estimates. Only estimates meeting the precision criterion (95% CI width < .80) were included. M = mean; SD = standard deviation.

Table S27

Two-Way ANOVA Results for Individual-Level OB Estimates by IAT Domain

IAT	Effect	*df*	*F*	*p*	η_p^2
Sexuality	Group Status (G)	1, 143915	440.27	< .001	.003
	Model Type (M)	1, 143915	1033.7	< .001	.007
	G × M	1, 143915	850.56	< .001	.006
Race	Group Status (G)	1, 567177	678.36	< .001	.001
	Model Type (M)	1, 567177	1415.5	< .001	.002
	G × M	1, 567177	9033.7	< .001	.016
Skin Tone	Group Status (G)	1, 128194	263.37	< .001	.002
	Model Type (M)	1, 128194	123.36	< .001	.001
	G × M	1, 128194	1187.2	< .001	.009
Age	Group Status (G)	1, 92737	186.61	< .001	.002
	Model Type (M)	1, 92737	725.61	< .001	.008
	G × M	1, 92737	184.80	< .001	.002
Disability	Group Status (G)	1, 68599	0.00	.966	.000
	Model Type (M)	1, 68599	45.35	< .001	.001
	G × M	1, 68599	373.49	< .001	.005
Weight	Group Status (G)	1, 103717	494.41	< .001	.005
	Model Type (M)	1, 103717	426.40	< .001	.004
	G × M	1, 103717	376.32	< .001	.004

Note. Results of 2 (Group Status: higher-status vs. lower-status) × 2 (Model Type: ingroup-affirming vs. outgroup-affirming) between-subjects ANOVAs on individual-level Overcoming Bias (OB) parameter estimates. Only estimates meeting the precision criterion (95% CI width < .80) were included. η_p^2 = partial eta-squared.

Table S28

Simple Effects of Model Type on Individual-Level OB Estimates by IAT Domain and Group Status

IAT	Group Status	*t*	*df*	*p*	*d*	95% CI
Sexuality	Higher-Status	18.01	46698.7	< .001	0.12	[0.11, 0.14]
	Lower-Status	39.64	15031.3	< .001	0.62	[0.58, 0.65]
Race	Higher-Status	-70.82	151087.8	< .001	-0.26	[-0.27, -0.25]
	Lower-Status	70.66	67314.3	< .001	0.54	[0.52, 0.55]
Skin Tone	Higher-Status	-7.56	31053.2	< .001	-0.06	[-0.08, -0.05]
	Lower-Status	35.01	23112.5	< .001	0.46	[0.43, 0.48]
Age	Higher-Status	9.52	14497.6	< .001	0.11	[0.09, 0.14]
	Lower-Status	26.49	12223.2	< .001	0.33	[0.31, 0.36]
Disability	Higher-Status	-13.41	14782.3	< .001	-0.16	[-0.18, -0.13]
	Lower-Status	13.22	3892.8	< .001	0.32	[0.27, 0.37]
Weight	Higher-Status	0.98	14510.9	.325	0.01	[-0.01, 0.03]
	Lower-Status	26.77	15718.0	< .001	0.31	[0.29, 0.34]

Note. Simple effects of model type (ingroup-affirming vs. outgroup-affirming) on Overcoming Bias (OB) parameter estimates, computed separately for each group status within each IAT domain. d = Cohen's d; 95% CI = 95% bootstrap confidence interval for Cohen's d. Positive d values indicate greater OB for ingroup-affirming than outgroup-affirming specifications. Only estimates meeting the precision criterion (95% CI width < .80) were included.

Age-Based Robustness Analyses

To assess the robustness of our findings to various age distribution cutoffs, we conducted two complementary sets of robustness analyses: an adult-only reanalysis excluding respondents under 18, and a sensitivity analysis testing alternative age brackets for the age IAT's status classification.

First, we re-ran the full set of primary analyses after excluding all respondents under 18 across all six IAT domains (adult-only N = 5,002,253, reduced from the full sample of N = 5,664,693; see Tables S29–S32). This reanalysis reproduced the pattern of results reported in the main text. All equality-constraint tests of OB were significant in the expected direction across both status groups and all six domains, with one exception: the OB effect among higher-status respondents in the disability domain was non-significant, matching the null result observed in the full sample (Table S29). The lower-status OB effect in the disability domain, non-significant in the full sample ($p = .247$), reached significance in the adult-only sample ($\Delta G^2 = 4.07$, $p = .044$). Tests of AC were significant across every domain and status group for both ingroup- and outgroup-target parameters (Tables S30–S31), and the binomial model-fit results (Table S32) likewise mirrored the full-sample pattern.

Next, we additionally tested whether the age IAT's findings depend on the specific age cutoffs used to define its higher- and lower-status groups by re-running the age-domain analyses under two alternative specifications: (1) restricting the lower-status ("older") category to ages 40–49, and (2) applying this same restriction together with an under-18 exclusion for the higher-status ("younger") category. Across both alternative specifications, the equality-constraint tests for OB and for AC (both ingroup- and outgroup-target) remained significant in the same direction as in the original analysis, with comparable effect sizes (Tables S33–S35). The OB

effects among both higher- and lower-status groups were once again significant across both specifications (all ps < .001).

Taken together, these analyses indicate that the effects reported in the main text are not an artifact of the age composition or age-bracket boundaries of the analytic sample.

Table S29

Overcoming Bias (OB) Equality Constraints, Filtered for Adult Respondents

IAT	Group Status	ΔG^2	df	*p*	Δw	w Ratio
Sexuality	Higher-Status	2622.7	1	< .001	.0004	1.01
	Lower-Status	11791.6	1	< .001	.0102	1.21
Race	Higher-Status	25150.1	1	< .001	.0022	1.11
	Lower-Status	29012.2	1	< .001	.0141	1.77
Skin Tone	Higher-Status	526.5	1	< .001	.0003	1.01
	Lower-Status	6764.4	1	< .001	.0124	1.97
Age	Higher-Status	236.5	1	< .001	.0005	1.05
	Lower-Status	1371.1	1	< .001	.0026	1.26
Disability	Higher-Status	0.0	1	1.000	.0000	1.00
	Lower-Status	4.1	1	.044	.0000	1.00
Weight	Higher-Status	25.6	1	< .001	.0000	1.00
	Lower-Status	4060.1	1	< .001	.0034	1.16

Note. Results of likelihood ratio tests comparing models with OB parameters freely estimated across ingroup-affirming and outgroup-affirming specifications versus models with OB constrained to equality. ΔG^2 = change in likelihood ratio chi-square; Δw = change in Cohen's w effect size; w Ratio = ratio of constrained to unconstrained w (values > 1 indicate worse fit when parameters are constrained to equality).

Table S30

Ingroup Association Activations (AC) Equality Constraints, Filtered for Adult Respondents

IAT	Group Status	ΔG^2	df	*p*	Δw	w Ratio
Sexuality	Higher-Status	80182.5	1	< .001	.0099	1.24
	Lower-Status	23153.1	1	< .001	.0173	1.36
Race	Higher-Status	39573.9	1	< .001	.0033	1.16
	Lower-Status	10906.0	1	< .001	.0065	1.35
Skin Tone	Higher-Status	17651.4	1	< .001	.0060	1.28
	Lower-Status	160.1	1	< .001	.0005	1.04
Age	Higher-Status	9262.8	1	< .001	.0130	2.23
	Lower-Status	6773.3	1	< .001	.0093	1.94
Disability	Higher-Status	20255.7	1	< .001	.0085	1.25
	Lower-Status	1358.0	1	< .001	.0028	1.08
Weight	Higher-Status	16018.2	1	< .001	.0112	1.64
	Lower-Status	5469.4	1	< .001	.0045	1.21

Note. Results of likelihood ratio tests comparing models with AC parameters for ingroup targets freely estimated across ingroup-affirming and outgroup-affirming specifications versus models with AC constrained to equality. ΔG^2 = change in likelihood ratio chi-square; Δw = change in Cohen's w effect size; w Ratio = ratio of constrained to unconstrained w (values > 1 indicate worse fit when parameters are constrained to equality).

Table S31

Outgroup Association Activations (AC) Equality Constraints, Filtered for Adult Respondents

IAT	Group Status	ΔG^2	df	*p*	Δw	w Ratio
Sexuality	Higher-Status	2625.7	1	< .001	.0004	1.01
	Lower-Status	101.1	1	< .001	-.0001	1.00
Race	Higher-Status	58791.1	1	< .001	.0048	1.23
	Lower-Status	10806.0	1	< .001	.0063	1.35
Skin Tone	Higher-Status	17012.8	1	< .001	.0061	1.29
	Lower-Status	420.4	1	< .001	.0012	1.09
Age	Higher-Status	8620.0	1	< .001	.0123	2.16
	Lower-Status	9211.0	1	< .001	.0119	2.20
Disability	Higher-Status	11838.5	1	< .001	.0053	1.16
	Lower-Status	3442.9	1	< .001	.0067	1.18
Weight	Higher-Status	15073.9	1	< .001	.0107	1.62
	Lower-Status	5425.9	1	< .001	.0045	1.21

Note. Results of likelihood ratio tests comparing models with AC parameters for outgroup targets freely estimated across ingroup-affirming and outgroup-affirming specifications versus models with AC constrained to equality. ΔG^2 = change in likelihood ratio chi-square; Δw = change in Cohen's w effect size; w Ratio = ratio of constrained to unconstrained w (values > 1 indicate worse fit when parameters are constrained to equality).

Table S32

Ingroup-Affirming Model Fits by IAT and Group Status, Filtered for Adult Respondents

IAT	Group Status	% Ingroup-Affirming	*p*
Sexuality	Higher-Status	71.84	< .001
	Lower-Status	58.84	< .001
Race	Higher-Status	73.06	< .001
	Lower-Status	55.16	< .001
Skin Tone	Higher-Status	74.03	< .001
	Lower-Status	47.59	< .001
Age	Higher-Status	73.22	< .001
	Lower-Status	24.31	< .001
Disability	Higher-Status	74.15	< .001
	Lower-Status	29.75	< .001
Weight	Higher-Status	75.45	< .001
	Lower-Status	29.20	< .001

Note. % Ingroup-Affirming = percentage of respondents whose data were better fit by the ingroup-affirming model specification (the complementary percentage were better fit by the outgroup-affirming specification). p-values are from exact binomial tests comparing the proportion of ingroup-affirming fits to chance (50%).

Table S33

Overcoming Bias (OB) Equality Constraint, Alternative Age-Bracket Specifications

Specification	Group Status	ΔG^2	*p*	Δw	w Ratio
Older Category Restricted to 40-49	Higher-Status	741.3	< .001	.0010	1.09
	Lower-Status	524.9	< .001	.0023	1.23
18 and Older, Older Category Restricted to 40-49	Higher-Status	236.5	< .001	.0005	1.05
	Lower-Status	524.9	< .001	.0023	1.23

Note. Results of likelihood ratio tests comparing models with OB parameters freely estimated across ingroup-affirming and outgroup-affirming specifications versus models with OB constrained to equality, for the age IAT under two alternative age-bracket specifications. ΔG^2 = change in likelihood ratio chi-square; Δw = change in Cohen's w effect size; w Ratio = ratio of constrained to unconstrained w (values > 1 indicate worse fit when parameters are constrained to equality).

Table S34

Ingroup Association Activation (AC) Equality Constraint, Alternative Age-Bracket Specifications

Specification	Group Status	ΔG^2	*p*	Δw	w Ratio
Older Category Restricted to 40-49	Higher-Status	15468.0	< .001	.0129	2.20
	Lower-Status	3311.0	< .001	.0101	2.00
18 and Older, Older Category Restricted to 40-49	Higher-Status	9262.8	< .001	.0130	2.23
	Lower-Status	3311.0	< .001	.0101	2.00

Note. Results of likelihood ratio tests comparing models with AC parameters for ingroup targets freely estimated across ingroup-affirming and outgroup-affirming specifications versus models with AC constrained to equality, for the age IAT under two alternative age-bracket specifications. ΔG^2 = change in likelihood ratio chi-square; Δw = change in Cohen's w effect size; w Ratio = ratio of constrained to unconstrained w (values > 1 indicate worse fit when parameters are constrained to equality).

Table S35

Outgroup Association Activation (AC) Equality Constraint, Alternative Age-Bracket Specifications

Specification	Group Status	ΔG^2	*p*	Δw	w Ratio
Older Category Restricted to 40-49	Higher-Status	13313.5	< .001	.0115	2.07
	Lower-Status	3711.3	< .001	.0112	2.11
18 and Older, Older Category Restricted to 40-49	Higher-Status	8620.0	< .001	.0123	2.16
	Lower-Status	3711.3	< .001	.0112	2.11

Note. Results of likelihood ratio tests comparing models with AC parameters for outgroup targets freely estimated across ingroup-affirming and outgroup-affirming specifications versus models with AC constrained to equality, for the age IAT under two alternative age-bracket specifications. ΔG^2 = change in likelihood ratio chi-square; Δw = change in Cohen's w effect size; w Ratio = ratio of constrained to unconstrained w (values > 1 indicate worse fit when parameters are constrained to equality).

Overcoming Bias as a Measure of Self-Regulatory Control

Within the QPM's formal structure, OB captures the probability that an activated evaluative association (AC) is successfully prevented from driving a response when it conflicts with the correctly detected task requirement (D). The model is agnostic about the precise operating principles OB reflects, and it is reasonable to ask whether the parameter reflects self-regulatory control or simply other processing and, if the former, what kind of control. We address these questions in turn.

Evidence for OB as a Measure of Self-Regulatory Control

Several converging lines of evidence validate OB as a measure of self-regulatory control. First, response deadlines reduced OB and D parameters but not AC and G, consistent with OB being a relatively resource-dependent, controlled process (2, 3). Second, OB declined systematically with age across a large lifespan sample, while the D parameter did not (8, 9). If OB merely reflected generic controlled processing, it should co-vary with D in this regard; the dissociation instead is consistent with the age-related deterioration of self-regulatory functioning (10–13). Third, alcohol consumption, whose primary cognitive implication is impaired response inhibition (14, 15), selectively reduced OB while leaving AC, D, and G largely intact (3), consistent with event-related potential (ERP) data from the same participants indicating selective impairment of inhibitory control (16).

Fourth, implementation intentions to respond positively when seeing a Black face increased OB but decreased D (6). This double dissociation between OB and D is consistent with the two parameters capturing distinct processes, and is consistent with OB being particularly sensitive to deliberate regulatory control strategies that D is not. Complementing this finding, counter-prejudicial training, via extensive practice mapping counter-stereotypical stimulus-response

pairings, reduced AC and enhanced D, while leaving OB unaffected (5). That is, practice seemingly strengthened participants' learned associations and their ability to identify correct responses, but not their capacity to override activated associations when those associations conflicted with correctly detected task requirements. Together with the implementation intentions finding, these two studies form a double dissociation: practiced task competence moves AC and D but not OB, whereas deliberate regulatory strategies move OB but not D, consistent with OB capturing self-regulatory control rather than learned accuracy-oriented stimulus discrimination.

Fifth, and related to the last point, Calanchini and colleagues (17) found that OB correlated strongly across IATs measuring conceptually unrelated attitude objects, whereas AC correlations declined systematically with decreasing conceptual overlap. This dissociation is consistent with OB reflecting a domain-general executive capacity rather than attitude-specific evaluative associations (18). Additionally, individual differences in internal and external motivations to control prejudice predict AC and D but not OB (19), further distinguishing the regulatory process captured by OB from motivationally driven accuracy monitoring. Finally, among participants with frequently activated anti-Muslim associations, higher OB predicted more favorable confederate-rated interaction quality with a Muslim confederate (20). This finding is consistent with OB capturing regulatory control that translates beyond the task itself, as participants who more successfully regulated their high levels of bias during the implicit measure were subsequently evaluated more positively during face-to-face interaction with presumably a target of that bias.

Considering Alternatives to OB as Strictly a Measure of Self-Regulatory Control

Alternatively, OB may simply reflect participants' choice to respond more cautiously: a conservative speed-accuracy tradeoff in which participants set a higher response threshold under anticipated conflict, producing fewer errors through more deliberate responding rather than through inhibition. A participant who checks responses more carefully on incongruent trials might show fewer bias-consistent errors, potentially inflating OB estimates without any genuine suppression or inhibition of biased responses occurring. The drift diffusion model (DDM; 21) provides a way to test this account, as its boundary separation parameter quantifies response threshold conservatism independently of other influences on responding. If OB and boundary separation capture much of the same process, controlling for the latter should eliminate OB effects. Gonsalkorale and colleagues (9) conducted exactly this test and found that OB differences between age groups survived boundary separation controls, indicating that OB captures variance in conflict resolution not reducible to differences in the extent of cautious responding enacted. This analysis was conducted on a different dataset than the present work, so it cannot rule out response caution as an account of the specific OB asymmetries observed here.

A related alternative is that OB reflects the detection of response conflict (i.e., conflict monitoring) rather than the implementation of inhibitory control. On this account, greater sensitivity to conflict on incongruent trials would increase the likelihood of withholding bias-consistent responses because heightened conflict detection more reliably triggers downstream control processes, even if OB does not capture the implementation of control. The ERP data reported by Bartholow et al. (16) speak to this question. Using a go/stop priming paradigm with alcohol administration, the authors found that the no-go N2, an ERP component indexing conflict detection (22, 23), was unaffected by alcohol and was inversely correlated with successful inhibition, suggesting that conflict was detected but could not be acted upon when

regulatory resources were impaired. In contrast, the negative slow wave (NSW), a component indexing implementation of cognitive control (24, 25), was selectively reduced by alcohol on trials requiring inhibition of prepotent responses. Its amplitude was linked to alcohol's effect on behavioral disinhibition. This neural double dissociation is consistent with OB operating at the level of implementing regulatory control rather than conflict monitoring. Taken together with the selective influences described above, the available evidence consistently positions OB as reflecting self-regulatory control that is unlikely to be reducible to cautious responding, conflict monitoring, automatic associations, or accuracy-oriented stimulus discrimination.

Distinguishing Types of Self-Regulatory Control: Proactive Versus Reactive Control

Research on cognitive control increasingly distinguishes between two broad modes of self-regulatory control (26). Proactive control operates anticipatorily: it involves the sustained maintenance of task-relevant goals in working memory prior to encountering potential conflict, functioning as an early-selection mechanism that can prevent interference before it arises. Reactive control operates differently: it is triggered by conflict itself, functioning as a late-correction mechanism that inhibits an already-activated competing representation on a trial-by-trial basis. Whether OB reflects proactive control, reactive control, or some combination of both is a critical question for understanding what the parameter does and does not capture. It is therefore useful to consider manipulations that differentially affect the two modes, which we consider below.

Alcohol Consumption

The most direct evidence linking OB to reactive rather than proactive control comes from research on the effects of alcohol consumption. The QPM was used to decompose behavioral data from Bartholow and colleagues' (16) race-stereotype priming task, in which participants

received doses of alcohol, and found that alcohol selectively reduced OB while leaving AC, D, and G largely intact (3). Crucially, Bartholow and colleagues (16) also recorded ERPs from those same participants, and found selective impairment of the negative slow wave (NSW), but not the no-go N2, both of which were time-locked to a randomly timed stop signal appearing on only 25% of trials. Because the stop signal was infrequent and unpredictable, it is unlikely that participants prepared inhibitory responses in advance, suggesting that the regulatory processing captured by the NSW specifically was reactive.

Converging evidence comes from experimental studies that directly manipulate alcohol dose and measure the two control modes independently. EEG theta-band analyses found that high-alcohol-dose intoxication decreased reactive control while paradoxically *increasing* proactive control as a compensatory response, suggesting reactive processing bears the primary burden of alcohol's inhibitory effects (15). An additional pair of studies finds that chronic use of alcohol impaired reactive inhibitory control while leaving proactive control intact (27). That proactive control was preserved while reactive control was selectively impaired strengthens the inference that OB's alcohol-sensitivity reflects reactive processing specifically.

Aging

The aging literature introduces tension to the definition of OB as reflecting purely reactive control. As discussed earlier, OB declined systematically across large lifespan samples (8, 9), a dissociation consistent with age-related deterioration of self-regulatory functioning (11, 12). However, research using proportion congruency manipulations to dissociate proactive and reactive control found that aging selectively impairs proactive control while reactive control remains relatively preserved across the adult lifespan (10, 13). If aging most consistently compromises proactive versus reactive control, then the age-related decline in OB aligns with an

account whereby the parameter reflects proactive control processing. This sits in tension with the alcohol and ERP evidence, and the current literature provides no clear basis for resolving it.

An Unresolved Picture of the Self-Regulatory Processes Reflected Under OB

Taken together, the existing literature strongly suggests that OB reflects a form of self-regulatory control but leaves unresolved whether that control operates proactively, reactively, or through a combination of both mechanisms. Evidence from alcohol administration and ERP studies most directly implicates reactive processing, whereas findings from the aging literature suggest the opposite. Resolving this tension will require studies that directly manipulate the two control modes within QPM paradigms. In the present work, we therefore interpret OB as reflecting self-regulatory control, functionally characterized as the prevention of activated associations from driving responses, without making stronger claims about whether this control operates proactively, reactively, or through both mechanisms.

The Quad Model Equations

The QPM is not fit to individual trials but to the aggregated frequencies of responses across 16 response categories, each with its own equation. These 16 categories are defined by crossing two factors: block arrangement and stimulus type. The IAT contains two block arrangements — one in which the higher-status group shares a response key with positive evaluative words and the lower-status group shares a key with negative evaluative words (e.g., *Young/Good* + *Elderly/Bad*), and one in which these pairings are reversed (e.g., *Young/Bad* + *Elderly/Good*). Within each block arrangement, responses are recorded separately for four stimulus types: higher-status group stimuli (e.g., young faces), lower-status group stimuli (e.g., elderly faces), positive evaluative words (e.g., *good*), and negative evaluative words (e.g., *bad*). This yields eight unique trial types, each producing a count of correct and incorrect responses, for a total of 16 response categories.

The equations for all 16 response categories are presented below for the higher-status model specification, which corresponds to the ingroup-affirming model for higher-status participants and the outgroup-affirming model for lower-status participants. The lower-status model specification follows the same structure with congruent and incongruent block assignments reversed.

Example QPM Equations

Tree 1: Young stimuli | Young/Good share a response key (correct = Good key)
t01 Correct: ACg + (1-ACg) × D + (1-ACg) × (1-D) × G
t02 Incorrect: (1-ACg) × (1-D) × (1-G)

Tree 2: Old stimuli | Old/Bad share a response key (correct = Bad key)
t03 Correct: ACb + (1-ACb) × D + (1-ACb) × (1-D) × (1-G)
t04 Incorrect: (1-ACb) × (1-D) × G

Tree 3: Good words | Young/Good share a response key (correct = Good key)
t05 Correct: ACg + (1-ACg) × D + (1-ACg) × (1-D) × G
t06 Incorrect: (1-ACg) × (1-D) × (1-G)

Tree 4: Bad words | Old/Bad share a response key (correct = Bad key)
t07 Correct: ACb + (1-ACb) × D + (1-ACb) × (1-D) × (1-G)
t08 Incorrect: (1-ACb) × (1-D) × G

Tree 5: Young stimuli | Young/Bad share a response key (correct = Bad key)
t09 Correct: ACg × D × OB + (1-ACg) × D + (1-ACg) × (1-D) × (1-G)
t10 Incorrect: ACg × D × (1-OB) + ACg × (1-D) + (1-ACg) × (1-D) × G

Tree 6: Old stimuli | Old/Good share a response key (correct = Good key)
t11 Correct: ACb × D × OB + (1-ACb) × D + (1-ACb) × (1-D) × G
t12 Incorrect: ACb × D × (1-OB) + ACb × (1-D) + (1-ACb) × (1-D) × (1-G)

Tree 7: Good words | Young/Bad share a response key (correct = Good key)
t13 Correct: (1-ACg) × D + (1-ACg) × (1-D) × G
t14 Incorrect: ACg + (1-ACg) × (1-D) × (1-G)

Tree 8: Bad words | Old/Good share a response key (correct = Bad key)
t15 Correct: (1-ACb) × D + (1-ACb) × (1-D) × (1-G)
t16 Incorrect: ACb + (1-ACb) × (1-D) × G

References

1. Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85(2), 197.
2. Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The Quadruple Process model of implicit attitudes. *Journal of Personality and Social Psychology*, 89(4), 469–487.
3. Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review*, 115(2), 314.
4. Beer, J. S., Stallen, M., Lombardo, M. V., Gonsalkorale, K., Cunningham, W. A., & Sherman, J. W. (2008). The Quadruple Process model approach to examining the neural underpinnings of prejudice. *NeuroImage*, *43*(4), 775-783.
5. Calanchini, J., Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2013). Counter-prejudicial training reduces activation of biased associations and enhances response monitoring. *European Journal of Social Psychology*, *43*(5), 321-325.
6. Calanchini, J., Lai, C. K., & Klauer, K. C. (2021). Reducing implicit racial preferences: III. A process-level examination of changes in implicit preferences. *Journal of Personality and Social Psychology*, *121*(4), 796.
7. Singmann, H., Heck, D. W., Barth, M., Erdfelder, E., Arnold, N. R., Aust, F., ... & Groß, J. (2024). Evaluating the robustness of parameter estimates in cognitive models: A meta-analytic review of multinomial processing tree models across the multiverse of estimation methods. *Psychological Bulletin*, 150(8), 965.

8. Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2009). Aging and prejudice: Diminished regulation of automatic race bias among older adults. *Journal of Experimental Social Psychology*, 45(2), 410-414.
9. Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2013) Measures of implicit attitudes may conceal differences in implicit associations. *Social Psychological and Personality Science*, 5, 271–278.
10. Ball, B. H., Peper, P., & Bugg, J. M. (2023). Dissociating proactive and reactive control in older adults. *Psychology and Aging*, *38*(4), 323.
11. Connelly, S. L., Hasher, L., & Zacks, R. T. (1991). Age and reading: the impact of distraction. *Psychology and aging*, *6*(4), 533.
12. Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *Psychology of learning and motivation*, *22*, 193-225.
13. Ileri-Tayar, M., Bugg, J. M., Dudey, T., & Braver, T. S. (2025). Proactive control declines while reactive control is preserved across the adult lifespan. *Journal of Experimental Psychology: General*.
14. Easdon, C. M., & Vogel-Sprott, M. (2000). Alcohol and behavioral control: impaired response inhibition and flexibility in social drinkers. *Experimental and clinical Psychopharmacology*, *8*(3), 387.
15. Stock, A. K., Wendiggensen, P., Ghin, F., & Beste, C. (2023). Alcohol-induced deficits in reactive control of response selection and inhibition are counteracted by a seemingly paradox increase in proactive control. *Scientific Reports*, *13*(1), 1097.

16. Bartholow, B. D., & Heinz, A. (2006). Alcohol and aggression without consumption: Alcohol cues, aggressive thoughts, and hostile perception bias. *Psychological Science*, 17(1), 30-37.
17. Calanchini, J., Sherman, J. W., Klauer, K. C., & Lai, C. K. (2014). Attitudinal and non-attitudinal components of IAT performance. *Personality and Social Psychology Bulletin*, *40*(10), 1285-1296.
18. Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of experimental psychology: General*, *133*(1), 101.
19. Gonsalkorale, K., Sherman, J. W., Allen, T. J., Klauer, K. C., & Amodio, D. M. (2011). Accounting for successful control of implicit racial bias: The roles of association activation, response monitoring, and overcoming bias. *Personality and Social Psychology Bulletin*, *37*(11), 1534-1545.
20. Gonsalkorale, K., von Hippel, W., Sherman, J. W., & Klauer, K. C. (2009). Bias and regulation of bias in intergroup interactions: Implicit attitudes toward Muslims and interaction quality. *Journal of Experimental Social Psychology*, *45*(1), 161-166.
21. Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59.
22. Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological review*, *108*(3), 624.
23. Nieuwenhuis, S., Yeung, N., Van Den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: effects of response conflict and trial type frequency. *Cognitive, affective, & behavioral neuroscience*, *3*(1), 17-26.

24. West, R., & Alain, C. (1999). Event-related neural activity associated with the Stroop task. *Cognitive Brain Research*, *8*(2), 157-164.

25. West, R., & Alain, C. (2000). Age-related decline in inhibitory control contributes to the increased Stroop effect observed in older adults. *Psychophysiology*, *37*(2), 179-189.

26. Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences, 16*(2), 106–113.

27. Baines, L., Field, M., Christiansen, P., & Jones, A. (2019). The effect of alcohol cue exposure and acute intoxication on inhibitory control processes and ad libitum alcohol consumption. *Psychopharmacology*, *236*(7), 2187-2199.