

Oppressed Groups Engender Implicit Positivity: Seven Demonstrations Using Novel and Familiar Targets



Benedek Kurdi¹, Amy R. Krosch², and Melissa J. Ferguson¹

¹Department of Psychology, Yale University, and ²Department of Psychology, Cornell University

Psychological Science
1–18

© The Author(s) 2023

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09567976231194588

www.psychologicalscience.org/PS



Abstract

Across seven preregistered studies in online adult volunteer samples ($N = 5,323$), we measured implicit evaluations of social groups following exposure to historical narratives about their oppression. Although the valence of such information is highly negative and its interpretation was left up to participants, implicit evaluations of oppressed groups shifted toward positivity, including in designs involving fictitious, well-known, and even self-relevant targets. The sole deviation from this pattern was observed in an experiment using a vignette about slavery in the United States, in response to which neither White nor Black Americans exhibited any change in implicit race attitudes. In line with propositional perspectives, these findings suggest that implicit evaluations (including, notably, implicit evaluations of well-known and self-relevant social groups) tend to change toward positivity in response to extremely negative information involving past oppression. However, macro-level phenomena, such as public awareness of histories of oppression, can modulate such updating processes.

Keywords

attitude change, Implicit Association Test, implicit evaluations, intergroup relations, oppression, open data, open materials, preregistered

Received 6/23/22; Revision accepted 7/26/23

Truthful discussions of past oppression are central to reconciliation following intergroup conflict (Gibson, 2004), racial socialization (Abaied et al., 2022), education about history in racially stratified societies (Southern Poverty Law Center, 2018), and societal responses following incidents revealing persistent effects of oppressive structures rooted in the past. For example, the 2020 murder of George Floyd by a Minneapolis police officer was followed by both considerable changes in public opinion (Reny & Newman, 2021) and an increased desire to learn about the history of anti-Black racism in the United States (Barrie, 2020).

However, remarkably little is known about the downstream consequences of encountering information about a group's past oppression for its present-day evaluations. Relevant empirical work has focused on explicit (self-reported) evaluations of victims of harm (Hafer, 2000; Lerner & Miller, 1978). In these studies, participants often report negative evaluations of victims

of unjust suffering, presumably to reduce cognitive dissonance. However, explicit victim derogation is not ubiquitous: Jordan and Kouchaki (2021) have provided evidence for positive shifts in explicit evaluations of individuals who become victims of harm.

Critically, even when explicit victim evaluations are positive, implicit (automatic) evaluations need not be, for at least two reasons. First, explicit victim evaluations cannot be taken at face value given social desirability concerns (Hafer, 2000; Lerner & Miller, 1978). Second, influential theoretical (Rydell & McConnell, 2006; Strack & Deutsch, 2004) and empirical (Gawronski et al., 2022) work has suggested that implicit evaluations reflect exclusively (or at least primarily) the sum total of

Corresponding Author:

Benedek Kurdi, University of Illinois Urbana-Champaign, Department of Psychology

Email: kurdi@illinois.edu

evaluative information encountered about a target. If this is the case, then although often considered a prerequisite for desirable outcomes, “consciousness raising exercises aimed at increasing awareness of social oppression may, ironically, strengthen automatic prejudices” (Uhlmann et al., 2006, p. 497). Thus, it is imperative to ask whether education about past oppression can create such unintended effects.

Notably, emerging evidence suggests that implicit evaluations can incorporate information going beyond simple stimulus pairings (Cone et al., 2017; De Houwer, 2014; Kurdi et al., 2022). For example, Zanon et al. (2014) found that implicit evaluations of nonwords (e.g., LOKANTA) reflected not merely the valence of the English words with which they had been paired (e.g., HAPPY) but also the relationship that the two stimuli shared. Specifically, when HAPPY was labeled as evaluatively equivalent to LOKANTA, LOKANTA became implicitly positive. However, this effect was attenuated when LOKANTA and HAPPY were described as opposite in meaning. Thus, implicit evaluations may also reflect the difference between perpetrators and victims, although both tend to appear in the context of highly negative events.

However, the present case differs from past tests in several critical ways. First, most past studies involve conditioning-like procedures with explicit instructions guiding participants’ trial-by-trial interpretation of stimulus pairings. By contrast, historical narratives do not come with express instructions to associate oppressed groups with the opposite of the narrative’s negative valence. Second, implicit evaluations are less likely to reflect relational information (e.g., perpetrator vs. victim) when the valence of the accompanying information is highly negative (Kurdi et al., 2022), which further raises the specter of ironic effects. Thus, Studies 1 and 2 provide a stringent proof-of-concept test of whether, contrary to associative ideas, highly negative oppression-related narratives can create implicit positivity toward novel groups.

Third, and critically, it is unclear whether relational effects on implicit evaluation generalize to well-known targets (Kurdi et al., 2022). Under certain associative accounts (Rydell & McConnell, 2006; Strack & Deutsch, 2004), well-established implicit evaluations should not change at all in response to minimal experimental interventions (see also Krosnick & Petty, 1995). And, even if they do, they should reflect the negative valence of oppression-related narratives rather than the oppressed groups’ status as victims rather than perpetrators. Driven by these considerations, Studies 3 to 5 examined implicit evaluations of well-known social groups following exposure to narratives about their oppression. In combination, then, the present studies speak to both

Statement of Relevance

Disseminating accurate information about past wrongdoing in intergroup contexts (including discrimination, slavery, or genocide) constitutes an indispensable first step toward reconciliation and restitution. However, worryingly, such information may produce ironic effects: Because oppression itself is extremely negative, implicit (automatic) evaluations of oppressed groups may shift in a negative, rather than positive, direction. Contrary to these ideas, we found that information about oppression changed implicit evaluations of social groups, including well-known and even personally relevant ones, toward positivity. The sole exception was a set of studies about slavery in the United States in which neither White nor Black Americans showed any change in implicit race attitudes. Together, these studies should alleviate worries about unintended evaluative effects of educating the public about past oppression. Moreover, they suggest that although information about oppression tends to create positive evaluations, macro-level phenomena (such as societal awareness of past wrongdoing) can affect learning in individual minds.

basic social cognitive processes and their applicability to issues of real-world inequality.

In Study 3, we used the Armenian genocide as a test case in a sample of Americans. This scenario was selected as an initial real-world test case because the groups involved were expected to be known, but not self-relevant, to participants. However, critically, a finding of implicit positivity toward (even familiar) oppressed groups need not generalize to cases involving self-relevant targets. After all, exposure to extremely negative information about an in-group, especially in the moral domain, can result in defensive responding, including out-group derogation (Branscombe et al., 1999; Ellemers et al., 2002). Thus, in Studies 4 and 5, we conducted exceedingly severe tests of implicit positivity toward oppressed groups among the aggressor group’s descendants: White Americans in the context of genocide against Native Americans (Study 4) and the enslavement of Black individuals (Study 5).

Studies 1 and 2: Historical Narratives About Novel Targets

We first probed implicit evaluations of novel groups following exposure to vignettes about their oppression.

Table 1. Overview of the Design of Studies 1 to 5

Study	Sample	Target category	Target label	Narrative	Learning condition		Control condition	Testing condition		
					Control	Experimental		V/A	V/C	A/C
Study 1	U.S.	Fictitious	Fictitious	Fictitious (Gregg et al., 2006)	×	✓		✓	✓	✓
Study 2	U.S.	Real	Fictitious	Armenian genocide	✓	✓	Control vignettes	✓	✓	✓
Study 3	U.S.	Real	Real	Armenian genocide	✓	✓	Control vignettes	✓	✓	✓
Study 4	White U.S.	Real	Real	Native American genocide (Rotella & Richeson, 2013)	✓	✓	No intervention	✓	×	×
Study 5A	White U.S.	Real	Real	Slavery	✓	✓	No intervention	✓	×	×
Study 5B	Non-U.S.	Real	Fictitious	Slavery	×	✓		✓	×	×
Study 5C	Black U.S.	Real	Real	Slavery	✓	✓	No intervention	✓	×	×

Note: V = Victim; A = Aggressor; C = Control.

We did so because historical narratives provide a particularly strong test of relational influences on implicit evaluation and because materials of this kind are often used to raise awareness of past oppression. In Study 1, the narrative was adapted from the work by Gregg et al. (2006); in Study 2, it was a historically accurate narrative about the Armenian genocide using fictitious group labels.

Method

Ethical approval. The project received ethical approval from the Institutional Review Board for Human Participant Research at Cornell University and from the Yale University Institutional Review Board.

Open science practices. We report all measures, manipulations, and exclusions in these and all remaining studies. The hypothesis, design, sample size, and participant exclusions were formally preregistered (<https://aspredicted.org/tr9c8.pdf> for Study 1A, <https://aspredicted.org/t5nj2.pdf> for Study 1B, and <https://aspredicted.org/gj4j4.pdf> for Study 2). All raw data files, analysis scripts, and materials used in these and all remaining studies are available via the Open Science Framework (<https://osf.io/cdftx/>).

The analyses reported below deviate from the preregistered analysis plan in the following ways. In Studies 1A and 1B, we did not exclude participants on the basis of their performance on the three manipulation check items given that the statistical inferences were identical with or without such exclusions. In Study 2 (as well as in Study 3 below), we used Bayesian mixed-effects models instead of frequentist linear models

(which were included in the preregistration because of a clerical error) because (a) Bayesian mixed-effects models have the ability to account for stimulus effects and (b) Study 1 used a Bayesian mixed-effects model, and we sought to maintain consistency of analytic approaches across studies. However, the use of frequentist linear models results in the same statistical inferences.

Participants and design. Participants were adult volunteers from the United States recruited from the Project Implicit educational website (<http://implicit.harvard.edu/>; $N = 914$ in Study 1 and $N = 1,209$ in Study 2). An overview of the design for this and all remaining studies is included in Table 1; details of the sample, reliabilities of the main dependent measures, and exploratory measures are provided in Table 2. Exploratory measures whose results are reported in the main text are additionally described in the relevant Method sections. Key demographic details by study are reported in Table 3.

In Studies 1A and 1B, for the purposes of the test phase, each participant was randomly assigned to one of three between-participants conditions: victim/aggressor ($n = 294$), aggressor/control ($n = 312$), and victim/control ($n = 308$). In Study 2, for the purposes of the learning phase, each participant was randomly assigned to a control condition ($n = 610$) or an experimental condition ($n = 599$). For the purposes of the test phase, similar to Study 1, each participant was assigned to one of three between-participants conditions: victim/aggressor ($n = 386$), aggressor/control ($n = 419$), and victim/control ($n = 404$). Participants were independently assigned to conditions in the learning phase and in the test phase.

Table 2. Information About Sample Size, Reliability of the Dependent Measures, and Exploratory Measures

Study	Sample size				Reliability of explicit evaluation measures			Reliability of implicit evaluation measures			Exploratory measure
	Total	Incomplete	Inattentive	Final	V	A	C	V/A	V/C	A/C	
Study 1A	475	7	5	463	.87	.97	.89	.69	.71	.73	Manipulation check, Interpersonal Reactivity Index (Davis, 1983)
Study 1B	477	19	7	451							
Study 2	1,428	28	39	1,209	.91	.95	.92	.70	.68	.68	Knowledge about Armenian genocide
Study 3	827	17	8	802	.92	.94	.92	.61	.63	.56	
Study 4	519	20	5	494				.67			24-item measure of explicit memory (Rotella & Richeson, 2013)
Study 5A	709	16	0	693	.91	.90		.65			
Study 5B	550	24	2	524	.91	.93		.67			Knowledge about slavery
Study 5C	718	21	10	687	.91	.92		.65			

Note: Participants who did not complete the Implicit Association Test (IAT) are included in the Incomplete column, and participants with a response latency of 300 ms or below on at least 10% of IAT trials, suggesting inattentive responding, are included in the Inattentive column. Reliability of the explicit evaluation measures was calculated using Cronbach's alpha, and reliability of the implicit evaluation measures was calculated on the basis of 500 random split halves. V = Victim; A = Aggressor; C = Control.

Procedure and measures. In the learning phase of Study 1, participants learned about three fictitious social groups via a historical narrative about their oppression and completed measures of implicit and explicit evaluation in the test phase (the latter were made optional).

Similar to Study 1, Study 2 consisted of a learning phase and a test phase, with some modifications. Crucially, in the learning phase, each participant was assigned to one of two conditions: a control condition or an experimental condition. In the control condition, they read short, evaluatively neutral vignettes about the climate of three regions (corresponding to Armenia, Turkey, and Portugal). In the experimental condition, they read a coherent narrative about the Armenian genocide. Importantly, in both conditions, fictitious labels were used to refer to the three groups. In the test phase, participants completed a measure of implicit evaluations, explicit evaluation items, and two items probing their preexisting knowledge about the Armenian genocide.

Learning phase. The learning phase in Study 1 involved a verbal narrative of about 550 words, which participants read at their own pace over six separate screens. The narrative was a shortened and adapted version of the narrative used by Gregg et al. (2006) with a neutral third group added. To ensure that participants read the vignette attentively, we did not allow them

to proceed to the next screen before 15 s or 20 s had elapsed (depending on the amount of text displayed on the particular screen). This feature of the procedure was maintained for all remaining studies.

At the beginning of the narrative, participants were introduced to three groups: the aggressor group, the victim group, and the control group. Three of four fictitious group labels (Bonnians, Jebbians, Laapians, and Niffians) were randomly selected to serve as group labels for the three groups. All three groups were described as living in geographic proximity to each other. However, it was additionally mentioned that a mountain range cut off the control group from any interaction with the two remaining groups. This detail was added to convey the idea (explicitly reiterated at the end of the narrative) that the control group was oblivious to, and did not have the opportunity to intervene in, the conflict between the aggressor group and the victim group.

The narrative characterized the aggressor group as "profoundly militaristic" and "highly prejudiced" against the victim group. By contrast, the victim group was described as "peaceable" and having "little interest in waging war." The crucial part of the narrative then offered a detailed account of the aggressor group waging an unprovoked military campaign against the victim group. Specifically, the aggressor group was described as having invaded the victim group's territory under

Table 3. Distribution of Key Demographic Variables (Studies 1–5)

Study	Nationality	Gender	Age	Race
Study 1	United States (100%)	Female (71%) Male (29%)	$M = 33$ years ($SD = 14$)	White (73%) Black (11%) Unknown (5%) Multiracial (5%) Others (6%)
Study 2	United States (100%)	Female (65%) Male (35%)	$M = 37$ years ($SD = 16$)	White (71%) Black (11%) Multiracial (7%) Others (12%)
Study 3	United States (100%)	Female (68%) Male (32%)	$M = 37$ years ($SD = 16$)	White (73%) Black (11%) Unknown (5%) Multiracial (5%) Others (6%)
Study 4	United States (100%)	Female (68%) Male (32%) Others (2%)	$M = 35$ years ($SD = 17$)	White (100%)
Study 5A	United States (100%)	Female (65%) Male (33%) Others (2%)	$M = 41$ years ($SD = 15$)	White (100%)
Study 5B	United Kingdom (23%) Canada (14%) Australia (10%) New Zealand (5%) India (5%) Others (43%)	Female (56%) Male (42%) Others (2%)	$M = 38$ years ($SD = 14$)	White (63%) Asian (19%) Hispanic (6%) Multiracial (5%) Others (8%)
Study 5C	United States (100%)	Female (68%) Male (30%) Others (2%)	$M = 37$ years ($SD = 15$)	Black (100%)

Note: Levels of categorical variables corresponding to less than 5% of the data have been collapsed into “Others.”

false pretenses, looted and burned settlements, killed a total of 83,000 people, and continued to massacre inhabitants even after the victim group had surrendered. At the end of the narrative, the episode was described as “one of the most shameful in the history of human conflict.”

In the experimental condition of Study 2, participants read a coherent narrative of about 500 words about the Armenian genocide. Specifically, participants were informed that the Turkish government was responsible for the killing of about 1.5 million Armenians between 1914 and 1922. (Group labels in both conditions were replaced with the same fictitious group labels used in Study 1. The real group labels are used here merely for ease of understanding.) It was mentioned that Turkish propaganda characterized the Armenians as a “fifth column” that sided with the Allied powers, the adversaries of the Ottoman Empire during World War I. The bulk of the narrative concentrated on the specific atrocities committed against the Armenians, including confiscation of their property, their forcible transfer to labor

battalions where they were killed or worked to death, death marches across the desert, and deportations to a network of 25 concentration camps.

To keep evaluations of the third group (corresponding to the Portuguese) neutral, similar to Study 1, the text mentioned that they were oblivious to the egregious events at the time, implying that they could not have intervened. To underscore the significance of the events described in the narrative, the text concluded by noting that when the details of the atrocities came to light, “both ordinary citizens and high-ranking diplomats all over the world expressed horror at what had been going on.”

In the control condition of Study 2 (additionally included to keep the design parallel to Study 3, which used the same design but real, rather than fictitious, group labels), participants read three short vignettes of about 170 words each on the climate of Armenia, Portugal, and Turkey in individually randomized order. The vignettes were found to be evaluatively neutral on a pretest and mentioned details such as temperature,

precipitation, and regional and seasonal variation in the weather.

The narrative used in Study 1, which was designed to address an unrelated theoretical question, has some features that may not be ideal for the present purposes. First, inconsistent with the idea of a historical narrative, the text does not specify the time at which the events occurred, which may inadvertently have created excessive psychological distance or decreased believability. Second, the text contains evaluatively relevant descriptions of the two main groups even before the conflict started unfolding, with one of them characterized as “militaristic” and the other one characterized as “peaceable.” Third, there is a clear economic status difference between the groups: It is mentioned that the aggressor group was forced to impose austerity measures on the population to finance military campaigns; by contrast, the victim group is described as living in “continuing economic prosperity.” Thus, implicit positivity toward the oppressed group may, at least in part, have been due to it being described as wealthy (Horwitz & Dovidio, 2016).

Critically, the historical narrative used in Study 2 (a) specified the historical period during which the events occurred (1914–1922), (b) did not contain detailed information about the groups prior to the start of the conflict, and (c) did not mention any economic status differences between the groups, to remove the peculiarities of the vignette used in Study 1.

Test phase. In the test phase, participants completed measures of implicit and explicit evaluation. In Study 1, the explicit evaluation items were optional, whereas in Study 2, they were mandatory.

Implicit evaluations were measured using a standard five-block Implicit Association Test (IAT; Greenwald et al., 1998). For each participant, two of the three targets from the learning phase were randomly selected to serve as targets on the IAT, resulting in three between-participants conditions: victim/aggressor, aggressor/control, and victim/control. The description of the IAT below uses the victim and aggressor targets as an example. The design was identical for the remaining two conditions, with the group labels and corresponding stimuli replaced.

In Block 1 of the IAT (20 trials; category practice), participants used the E and I keys to sort stimuli corresponding to the victim and aggressor groups used in the learning phase. Category stimuli were five versions of the group label for each group: (a) singular capitalized (e.g., “Laapian”), (b) plural capitalized (e.g., “Laapians”), (c) singular all lowercase (e.g., “laapian”), (d) singular all caps (e.g., “LAAPIAN”), and (e) plural all caps (e.g., “LAAPIANS”). In addition, a font randomly

selected from the following five fonts was used on each trial: (a) purple, 3-em font size, bold; (b) green, 2.5-em font size, italicized; (c) maroon, 4-em font size, serif; (d) yellow, 3.5-em font size, cursive; and (e) blue, 2-em font size. These variations were created to ensure that participants were unable to categorize stimuli on the basis of perceptual features alone. The corresponding group names were used as category labels.

In Block 2 (20 trials; attribute practice), participants sorted the positively and negatively valenced adjectives described above using the same keys. The words “good” and “bad” served as attribute labels. In Block 3 (40 trials; first combined block), participants used one key to sort names from the victim group and positive adjectives and a different key to sort names from the aggressor group and negative adjectives (or vice versa). In Block 4 (20 trials; reversed category practice), participants sorted the names belonging to the victim and aggressor groups anew, with the placement of the two groups reversed. That is, if in Blocks 1 to 3, names from the victim group were sorted using the E key and names from the aggressor group using the I key, then in Block 4, the E key was used for the aggressor group and the I key was used for the victim group. In Block 5 (40 trials; second combined block), participants completed the same type of sorting task as in Block 3 but with the mapping between groups and valences reversed. That is, if in Block 3, names from the victim group were sorted together with positive adjectives and names from the aggressor group were sorted together with negative adjectives, then Block 5 required participants to sort aggressor together with positive and victim together with negative.

Responding on the IAT was scored using the improved scoring algorithm (Greenwald et al., 2003), resulting in an IAT D score expressing the relative implicit preference for the victim group over the aggressor group, the aggressor group over the control group, and the victim group over the control group, respectively.

In Study 1, participants used 7-point Likert-type scales to rate each of the three groups on the positive and negative adjectives used as attribute stimuli on the IAT. Each of the three groups was rated on a separate screen, with the order of the three screens and the order of the adjectives within each screen individually randomized. To parallel the IAT D scores, we calculated three explicit evaluation difference scores: victim/aggressor, aggressor/control, and victim/control.

In Study 2, to shorten the procedure, we had participants complete explicit evaluation items only for the two groups featured on the IAT (rather than all three groups, as in Study 1). Moreover, explicit evaluations were measured using 100-point slider scales rather than 7-point Likert-type scales.

Participants in Study 2 completed two explicit items measuring their knowledge of the Armenian genocide. First, using an open-ended item, we asked them to guess the three real-world groups/countries that the study had been about. Second, they were asked to indicate on a 100-point slider scale how knowledgeable they were about the Armenian genocide prior to the study. These items were collected for exploratory purposes.

Analytic strategy. Condition effects on implicit evaluations were investigated using Bayesian mixed-effects models with random intercepts for group names, implemented in the *brms* package in the R programming environment (Bürkner, 2017; Version 4.2.1, R Core Team, 2022). Unless otherwise noted, all models used default priors. Marginal means with their corresponding highest density intervals (HDIs) were obtained using the *emmeans* package (Lenth, 2020). Statistical inferences relied on the idea of indirect testing: If the 95% HDI did not overlap with zero, we inferred the presence of a significant effect; if it did, we refrained from making such inferences.

Because participants completed multiple measures of explicit evaluation, the Bayesian mixed-effects model with explicit evaluations as the dependent measure additionally included random intercepts for participants. For ease of interpretation, standardized (but not centered) versions of the dependent variables were entered into the models. Thus, regression coefficients in this and all remaining studies can be interpreted as standardized effect sizes.

Given that Study 1A yielded inconclusive results, a second round of data collection (Study 1B) was completed. To combine data from the two studies in a principled way, as preregistered, we used the regression coefficients obtained in Study 1A as informative priors in the Bayesian models used to analyze data from Study 1B and report them below (see Kruschke, 2010). The analytic strategy for Study 2 was the same as in Study 1, with two exceptions. First, given that no previous data collection using the same design was available, uninformative (default) priors were used. Second, because only one explicit evaluation difference score was available for each participant, no random intercepts for participants were added to the model with explicit evaluations as the dependent variable.

Results

Study 1.

Descriptive statistics. Figure 1 shows the distribution of explicit and implicit evaluations by condition for Study 1, along with condition means and 95% HDIs. Explicit and implicit evaluations showed similar patterns. Specifically,

participants exhibited an explicit ($M = 2.88$, $SD = 2.96$) and implicit ($M = 0.13$, $SD = 0.46$) preference for the victim over the aggressor group, an explicit ($M = -1.87$, $SD = 2.08$) and implicit ($M = -0.10$, $SD = 0.48$) preference for the control over the aggressor group, and crucially, an explicit ($M = 1.01$, $SD = 2.01$) and implicit ($M = 0.06$, $SD = 0.48$) preference for the victim over the control group.

Explicit evaluations. The marginal means derived from the Bayesian mixed-effects model fit to the data from Study 1B with informative priors derived from Study 1A confirmed the pattern of descriptive statistics reported above. Specifically, participants exhibited significant explicit preferences for the victim over the aggressor group, $\beta = 0.80$, 95% HDI = [0.70, 0.91]; the control over the aggressor group, $\beta = -0.57$, 95% HDI = [-0.68, -0.46]; and the victim over the control group, $\beta = 0.56$, 95% HDI = [0.46, 0.67].

Implicit evaluations. The same pattern of results was obtained on the theoretically crucial measure of implicit evaluation, with a significant preference for the victim over the aggressor group, $\beta = 0.32$, 95% HDI = [0.17, 0.46]; the control over the aggressor group, $\beta = -0.23$, 95% HDI = [-0.38, -0.08]; and the victim over the control group, $\beta = 0.18$, 95% HDI = [0.03, 0.32].

Study 2.

Descriptive statistics. Figure 2 shows the distribution of explicit and implicit evaluations by condition for Study 2, along with condition means and 95% HDIs. As expected given the nature of the materials and random assignment, explicit and implicit evaluations did not differ from neutrality in the control condition for any comparison, including victim/aggressor ($M = -2.81$, $SD = 18.57$ for explicit evaluations and $M = 0.03$, $SD = 0.45$ for implicit evaluations), aggressor/control ($M = 1.61$, $SD = 28.11$ for explicit evaluations and $M = -0.03$, $SD = 0.43$ for implicit evaluations), and victim/control ($M = 0.98$, $SD = 32.13$ for explicit evaluations and $M = 0.02$, $SD = 0.45$ for implicit evaluations).

In the experimental condition, the pattern of results was highly similar to the one obtained in Study 1: Participants exhibited an explicit ($M = 62.39$, $SD = 63.57$) and implicit ($M = 0.20$, $SD = 0.47$) preference for the victim over the aggressor group, an explicit ($M = -53.90$, $SD = 69.41$) and implicit ($M = -0.14$, $SD = 0.48$) preference for the control over the aggressor group, and crucially, an explicit ($M = 12.91$, $SD = 64.88$) and implicit ($M = 0.11$, $SD = 0.45$) preference for the victim over the control group.

Explicit evaluations. Adjusting for baseline evaluations measured in the control condition, we found that participants in the experimental condition exhibited significant explicit preferences for the victim over the

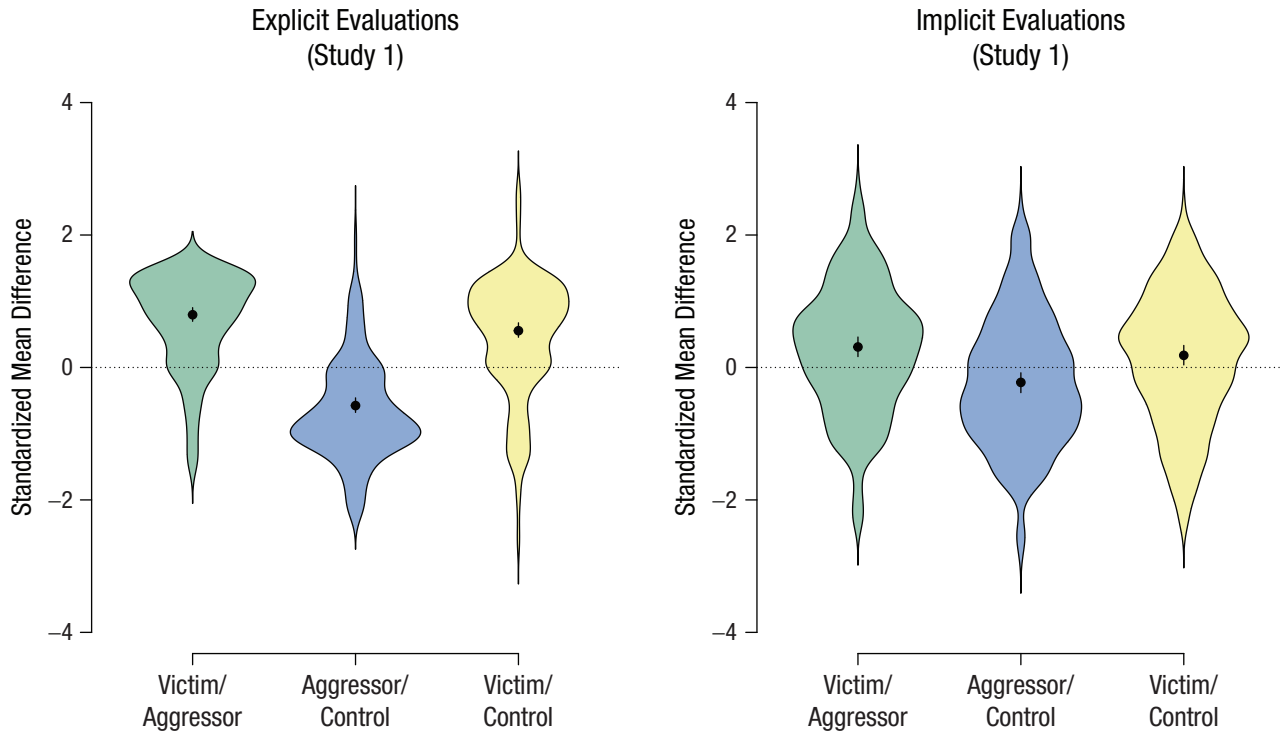


Fig. 1. Distribution of explicit and implicit evaluations by condition (Study 1), displayed in standardized units to ensure comparability. The dashed horizontal line marks neutrality, and the solid dots show condition means. Error bars represent 95% highest density intervals. Positive scores indicate preference for the victim over the aggressor group, the aggressor over the control group, and the victim over the control group, respectively.

aggressor group, $\beta = 1.08$, 95% HDI = [0.91, 1.26]; the control over the aggressor group, $\beta = -0.91$, 95% HDI = [-1.10, -0.76]; and the victim over the control group, $\beta = 0.20$, 95% HDI = [0.03, 0.36].

Implicit evaluations. Similarly, adjusting for baseline evaluations measured in the control condition, we found that participants in the experimental condition exhibited an implicit preference for the victim over the aggressor group, $\beta = 0.37$, 95% HDI = [0.18, 0.56]; the control over the aggressor group, $\beta = -0.24$, 95% HDI = [-0.42, -0.06]; and the victim over the control group, $\beta = 0.17$, 95% HDI = [-0.01, 0.36]. However, it should be noted that the latter HDI slightly overlaps with zero (corresponding to a p value of .054 in the frequentist analysis).

Explicit knowledge. In exploratory analyses, we investigated any effects of participants' preexisting knowledge of the Armenian genocide on the results of Study 2. Notably, only 52 out of 457 participants in the experimental condition (corresponding to 11%) mentioned the words Armenian or Turkish in their open-ended text response. The most frequent responses to this item referred to Jews as the victim group and/or Germans or Nazis as

the aggressor group ($n = 305$ or 67%). Thus, given the small number of participants with a correct response, no further analyses involving this variable were possible. Accordingly, participants' self-reported knowledge of the Armenian genocide was significantly below the midpoint of the scale ($M = -28.80$, $SD = 24.54$), 95% HDI = [-30.15, -27.31]. Self-reported knowledge was not associated with implicit or explicit evaluations.

Discussion

Implicit evaluations of oppressed social groups were mildly positive following exposure to two different historical narratives, both using extremely negative and vivid language. At the same time, the perpetrators of oppression were subject to highly negative implicit evaluations. The fact that implicit evaluations shifted toward positivity in response to highly negatively valenced materials suggests that they can incorporate rich sources of information going beyond simple stimulus associations. Notably, such learning emerged without any express instructions to mentally reverse the valence of the materials to which participants were exposed.

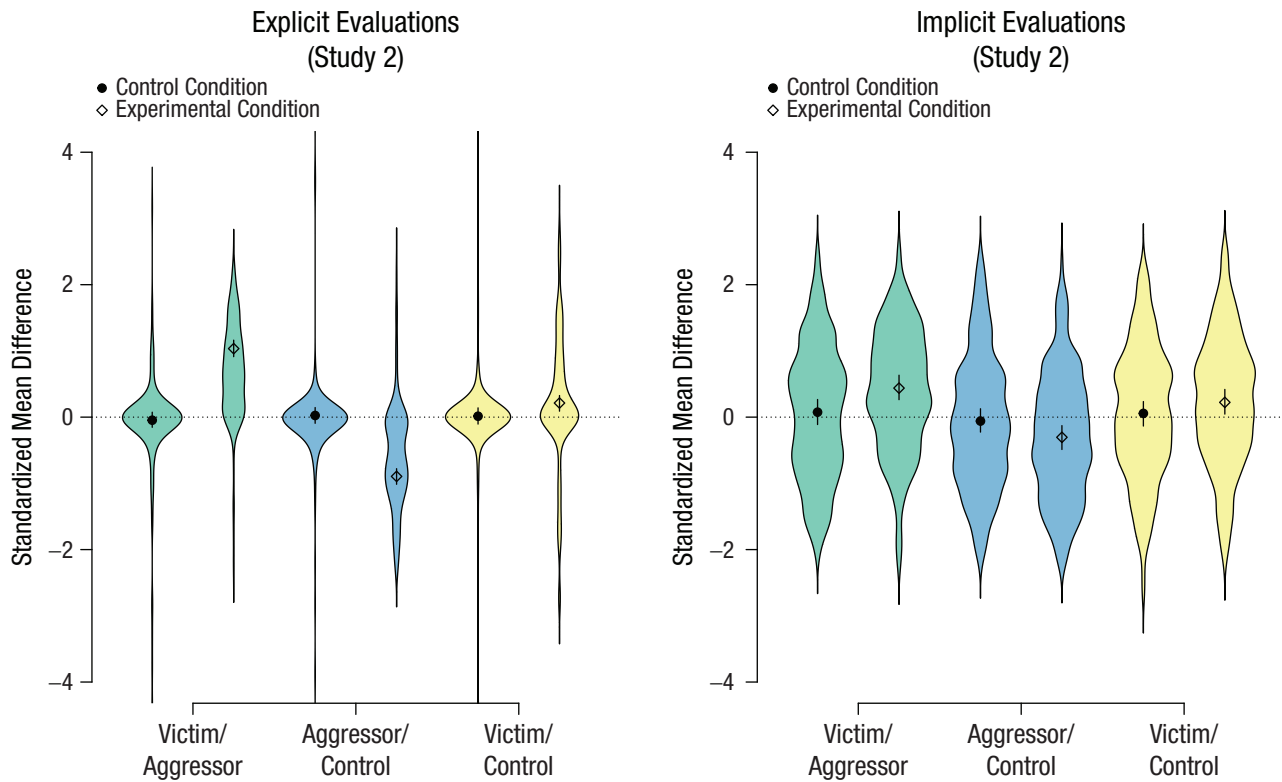


Fig. 2. Distribution of explicit and implicit evaluations by condition (Study 2), displayed in standardized units to ensure comparability. The dashed horizontal line marks neutrality. Solid dots denote means in the control condition, and open diamonds denote means in the experimental condition. Error bars represent 95% highest density intervals. Positive scores indicate preference for the victim over the aggressor group, the aggressor over the control group, and the victim over the control group, respectively.

Studies 3 to 5: Historical Narratives About Familiar and Self-Relevant Targets

The results of Studies 1 and 2 are instructive regarding the basic cognitive processes undergirding implicit social evaluation. However, it is important to probe whether similar updating processes sensitive to relational content emerge in the context of preexisting social targets that are known and even self-relevant to participants. Such tests have the potential to provide constraints on theorizing about social cognitive processes and are directly informative with respect to the evaluative consequences of awareness raising focused on past oppression.

Notably, relevant tests involving familiar targets are almost entirely missing from the literature and, in the rare cases that they have been conducted, have produced mixed results at best (Kurdi et al., 2022). Indeed, the results obtained using fictitious groups may not generalize to well-known targets because well-rehearsed attitudes are more difficult to change than newly established ones (Krosnick & Petty, 1995). Moreover, it is conceivable that identity-related motives may interfere

with such learning, especially when one's in-group members are portrayed as having committed egregiously immoral acts (Branscombe et al., 1999; Ellemers et al., 2002).

Method

Preregistration. The hypotheses, design, sample size, and participant exclusions were formally preregistered (<https://aspredicted.org/dr27p.pdf> for Study 3, <https://aspredicted.org/ze8e9.pdf> for Study 4, <https://aspredicted.org/u2xb4.pdf> for Study 5A, <https://aspredicted.org/838ct.pdf> for Study 5B, and <https://aspredicted.org/9dv9v.pdf> for Study 5C).

Participants and design. Participants were adult volunteers from the Project Implicit educational website in all five studies. In Study 3, 802 participants were recruited from the United States without respect to their race; in Studies 4 and 5A, participants were exclusively White individuals from the United States ($N_s = 693$ and 494); in Study 5B, 524 participants were recruited from outside the United States, with 81 unique nationalities represented, the most

common being the United Kingdom ($n = 115$), Canada ($n = 70$), Australia ($n = 48$), New Zealand ($n = 25$), and India ($n = 23$); and in Study 5C, participants were exclusively Black individuals from the United States ($N = 687$).

For the purposes of the learning phase, each participant in Studies 3, 4, 5A, and 5C was randomly assigned to a control condition ($n = 398$ in Study 3, $n = 267$ in Study 4, $n = 352$ in Study 5A, and $n = 334$ in Study 5C) or an experimental condition ($n = 404$ in Study 3, $n = 227$ in Study 4, $n = 341$ in Study 5A, and $n = 353$ in Study 5C). All participants in Study 5B completed the experimental condition given that Study 5B involved targets referred to using fictitious group labels. Thus, given random assignment to group labels, evaluations are interpretable relative to neutrality even in the absence of a separate control condition.

For the purposes of the test phase, each participant in Study 3 was assigned to one of three between-participants conditions: victim/aggressor ($n = 263$), aggressor/control ($n = 275$), and victim/control ($n = 264$). Participants were assigned to conditions in the learning and test phases independently of each other. In Studies 4 and 5, all participants completed IATs comparing the victim group with the aggressor group.

Procedure and measures. In all five studies, participants in the experimental condition completed a learning phase in which they learned about the historical oppression of a target group and then completed measures of implicit and explicit evaluation.

Learning phase. In Study 3, the target groups were Armenians, Turks, and Portuguese. The procedure and measures in Study 3 were identical to those used in Study 2, with the crucial exception that groups were referred to using historically accurate, rather than fictitious, labels. Thus, this study tested whether preexisting familiarity with the targets modulated the results of Study 2.

In the learning phase of Study 4, each participant was assigned to one of two conditions: a control condition or an experimental condition. In the control condition, they proceeded directly to the test phase of the experiment. In the experimental condition, they read a narrative about the mistreatment of the Illiniwek Native Americans by White Americans. This narrative was slightly adapted from a similar narrative used in a set of studies by Rotella and Richeson (2013). The goal of this study was to examine whether self-relevant narratives with the participant's in-group responsible for group-based violence can produce updating in line with the patterns observed in the remaining studies.

The procedure and measures in Studies 5A and 5C were similar to those in Study 4, with the exception that in the experimental condition, participants read a

historical narrative of approximately 450 words about the history of slavery in the United States. The vignette was adapted from the relevant Wikipedia article and discussed, among other topics, the legal institution of slavery, the mistreatment of enslaved Black people, slave auctions, and sexual abuse. Participants in the control condition proceeded directly to the dependent measures. The purpose of these studies was to probe whether the self-relevant nature of the targets modulated evaluative learning about oppressed social groups among White Americans (Study 5A) and Black Americans (Study 5C).

In Study 5B, all participants completed the same experimental manipulation involving a version of the historical narrative used in Study 5A. However, crucially, the participants were not from the United States and fictitious group labels were used. Specifically, group labels were randomly selected from Bonnians, Jebbians, Laapians, and Niffians (see Studies 1 and 2). Accordingly, this study constitutes a test of whether the materials used in Studies 5A and 5C are intrinsically different from the materials used in Studies 1 to 4, or whether the self-relevant nature of the vignette interfered with learning in Studies 5A and 5C (which, unlike Studies 1–4, produced no movement in the direction of implicit positivity toward the oppressed group). Study 5B used fictitious group labels to which participants were randomly assigned. Thus, given that deviations from zero are interpretable as evaluative preferences, no control condition was included.

Test phase. In Studies 3 and 5B, similar to Studies 1 to 2, the IAT category labels were the group labels, and category stimuli were the group labels printed in different fonts; in Study 4, the category labels were “White Americans” and “Native Americans,” and category stimuli were six family names randomly selected from the set Adams, Allen, Baker, Clark, Hall, Nelson, Scott, and Wright for White Americans and Awiakta, Wahchumwah, Chippewa, Suwake, Tsosie, Akiwenzie, Ojibway, Pawaush, Apache, Chosa, Kiatta, Homma, Pappan, and Yerxa for Native Americans; and in Studies 5A and 5C, the IAT category labels were “White Americans” and “Black Americans,” and category stimuli were grayscale photographs of six category members each (three women and three men). The IAT attribute labels and stimuli were the same as in Studies 1 and 2, and in Study 4 they also additionally included “excellent” and “perfect” for the good attribute and “nasty” and “vile” for the bad attribute.

In Studies 3 and 5, explicit evaluation items were identical to the ones used in Study 2. In Study 4, to shorten the procedure, explicit evaluations of the target groups were measured using one 100-point feeling thermometer item each (“How warmly or coldly do you

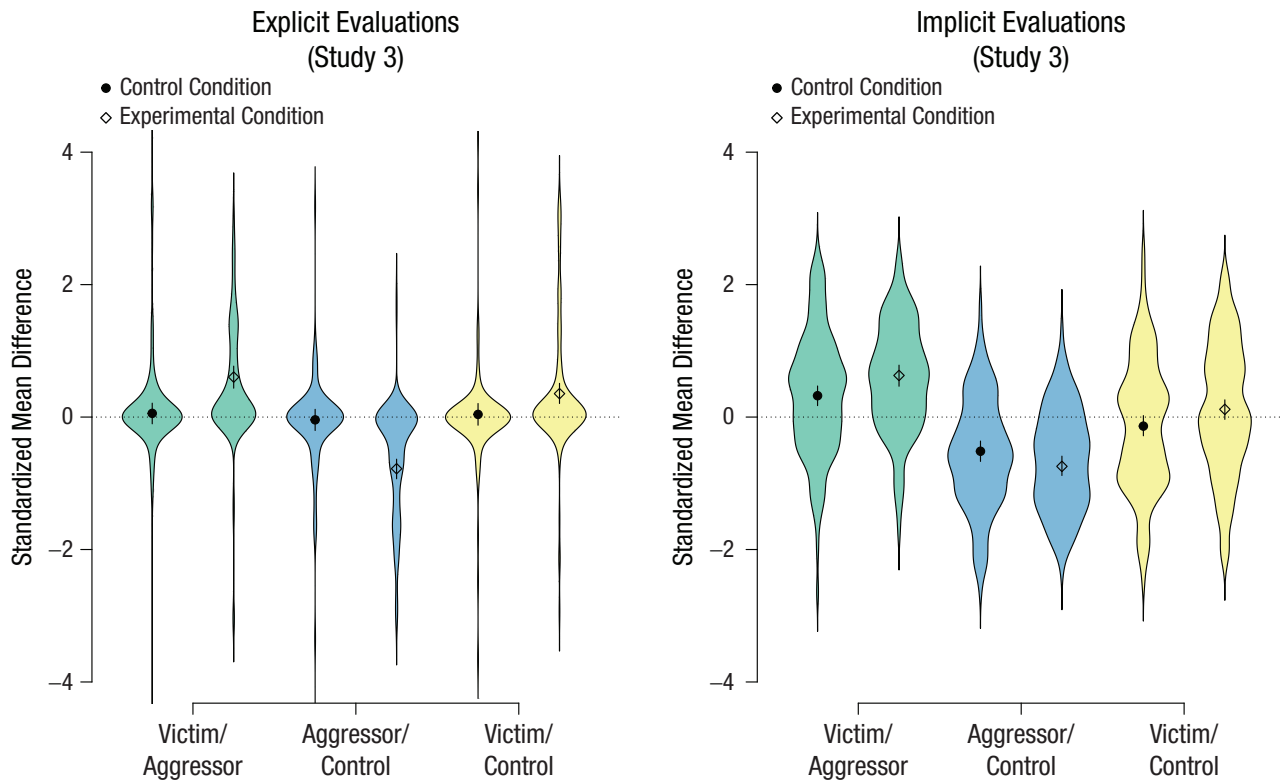


Fig. 3. Distribution of explicit and implicit evaluations by condition (Study 3), displayed in standardized units to ensure comparability. The dashed horizontal line marks neutrality. Solid dots denote means in the control condition, and open diamonds denote means in the experimental condition. Error bars represent 95% highest density intervals. Positive scores indicate preference for the victim over the aggressor group, the aggressor over the control group, and the victim over the control group, respectively.

feel toward Native/White Americans?”), with the endpoints labeled “Extremely coldly” and “Extremely warmly,” respectively. The two items were administered in random order.

In addition, at the end of Study 3, participants were asked (a) to report whether they assumed the study would be about the Armenian genocide after the protagonists had been introduced and (b) to estimate the extent of their knowledge about the Armenian genocide using a 100-point scale. At the end of Study 5B, participants were asked (a) to guess which real-world groups the story was about and (b) to estimate the extent of their knowledge about slavery in the United States on a 100-point scale.

Results

Study 3: Armenian genocide (U.S. sample).

Descriptive statistics. Figure 3 shows the distribution of explicit and implicit evaluations by condition, along with condition means and 95% HDIs. The pattern of explicit evaluations was identical to the one observed in Study 2, which used the same materials but fictitious group labels. Specifically, explicit evaluations did not

deviate from neutrality in the control condition, including for the victim/aggressor ($M = 2.49$, $SD = 25.37$), aggressor/control ($M = -1.85$, $SD = 12.00$), and victim/control ($M = 1.83$, $SD = 27.23$) comparisons. The expected shifts were observed in the experimental condition, with an explicit preference for the victim over the aggressor group ($M = 27.29$, $SD = 57.78$), an explicit preference for the control over the aggressor group ($M = -35.13$, $SD = 52.70$), and an explicit preference for the victim over the control group ($M = 15.87$, $SD = 46.38$).

Unlike in Study 2, the measure of implicit evaluation exhibited deviations from neutrality even in the control condition, with an implicit preference for the victim over the aggressor group ($M = 0.14$, $SD = 0.39$) and the control over the aggressor group ($M = -0.23$, $SD = 0.35$). The victim and the control groups were found to be evaluatively equivalent ($M = -0.06$, $SD = 0.41$). Critically, in spite of these deviations from neutrality at baseline, the pattern of changes in implicit evaluations as a result of the experimental manipulation was in line with Study 2: Means in the experimental condition reflected an increase in implicit positivity toward the victim relative to the aggressor group ($M = 0.28$, $SD = 0.39$), a sizeable increase in implicit negativity toward the aggressor

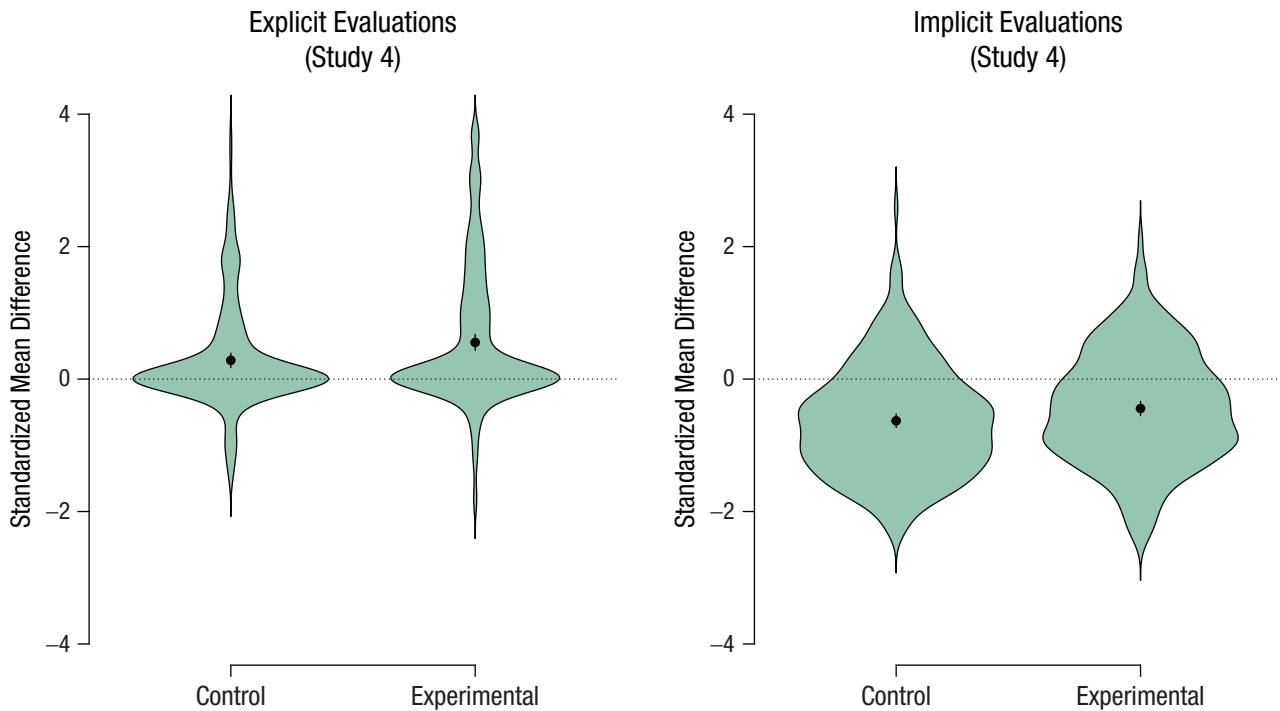


Fig. 4. Distribution of explicit and implicit evaluations by condition (Study 4), displayed in standardized units to ensure comparability. The dashed horizontal line marks neutrality. Solid dots denote condition means. Error bars represent 95% highest density intervals. Positive scores indicate preference for Native Americans (the victim group) over White Americans (the aggressor group).

relative to the control group ($M = -0.33$, $SD = 0.38$), and crucially, increased implicit positivity toward the victim relative to the control group ($M = 0.05$, $SD = 0.43$).

Explicit evaluations. The marginal means derived from the Bayesian mixed-effects model fit to the data confirmed the pattern of descriptive statistics described above. Specifically, adjusting for baseline evaluations measured in the control condition, we found that participants in the experimental condition exhibited significant explicit preferences for the victim over the aggressor group, $\beta = 0.55$, 95% HDI = [0.32, 0.76]; the control over the aggressor group, $\beta = -0.74$, 95% HDI = [-0.94, -0.51]; and the victim over the control group, $\beta = 0.31$, 95% HDI = [0.08, 0.53].

Implicit evaluations. The same pattern of results was obtained on the theoretically crucial measure of implicit evaluation. Specifically, adjusting for baseline evaluations measured in the control condition, we found that participants in the experimental condition exhibited an implicit preference for the victim over the aggressor group, $\beta = 0.30$, 95% HDI = [0.07, 0.50]; an implicit preference for the control over the aggressor group, $\beta = -0.22$, 95% HDI = [-0.41, -0.01]; and an implicit preference for the victim over the control group, $\beta = 0.25$, 95% HDI = [0.03, 0.46].

Explicit knowledge. Most participants ($n = 447$ or 58%) reported not assuming that the study would be about the Armenian genocide. Accordingly, participants' self-reported knowledge of the Armenian genocide was significantly below the midpoint of the scale ($M = -28.10$, $SD = 25.87$), 95% HDI = [-29.79, -26.14]. Neither of these variables was related to explicit or implicit evaluations.

Study 4: genocide against Native Americans (White U.S. sample).

Descriptive statistics. Figure 4 shows the distribution of explicit and implicit evaluations by condition, along with condition means and 95% HDIs. On the measure of explicit evaluation, participants expressed an out-group preference in both conditions, likely indicative of self-presentation concerns. This out-group preference was stronger in the experimental condition ($M = 15.03$, $SD = 28.10$) than in the control condition ($M = 7.70$, $SD = 21.12$). On the measure of implicit evaluation, participants exhibited an in-group preference overall. However, the condition effect was similar to that observed on the measure of explicit evaluation: The in-group preference was attenuated in the experimental condition ($M = -0.23$, $SD = 0.44$) relative to the control condition ($M = -0.33$, $SD = 0.43$).

Explicit evaluations. The marginal means derived from the Bayesian linear model fit to the data confirmed

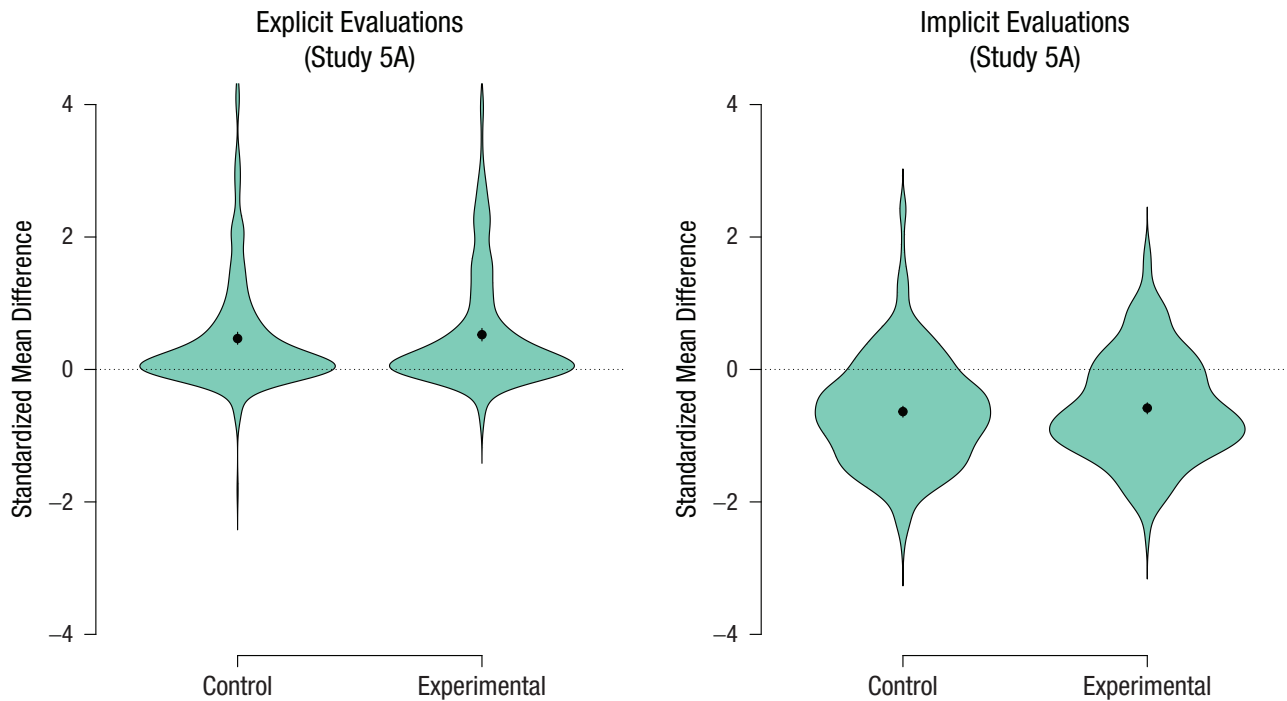


Fig. 5. Distribution of explicit and implicit evaluations by condition (Study 5A), displayed in standardized units to ensure comparability. The dashed horizontal line marks neutrality. Solid dots denote condition means. Error bars represent 95% highest density intervals. Positive scores indicate preference for Black Americans (the victim group) over White Americans (the aggressor group).

the pattern of descriptive statistics described above. Specifically, explicit out-group preference was stronger in the experimental condition, $\beta = 0.56$, 95% HDI = [0.44, 0.68], than in the control condition, $\beta = 0.28$, 95% HDI = [0.18, 0.39], resulting in a significant condition difference in line with Studies 1 to 3, $\beta = 0.27$, 95% HDI = [0.11, 0.44].

Implicit evaluations. A similar pattern of results was obtained on the theoretically crucial measure of implicit evaluation. Specifically, implicit in-group preference was attenuated in the experimental condition, $\beta = -0.45$, 95% HDI = [-0.56, -0.35], relative to the control condition, $\beta = -0.63$, 95% HDI = [-0.73, -0.53], resulting in a significant condition difference in line with the previous studies, $\beta = 0.19$, 95% HDI = [0.04, 0.33].

Study 5A: slavery (White U.S. sample).

Descriptive statistics. Figure 5 shows the distribution of explicit and implicit evaluations by condition for Study 5A, along with condition means and 95% HDIs. On the measure of explicit evaluation, participants expressed an out-group preference in both conditions, suggesting that—similar to Study 4—their responses were influenced by self-presentation concerns. This out-group preference did not differ by condition ($M = 14.60$, $SD = 27.79$ in the control condition and $M = 16.35$, $SD = 26.35$ in the experimental condition). On the measure of implicit evaluation, participants exhibited an in-group preference

overall. Unlike in previous studies, and similar to the measure of explicit evaluation, we observed no condition differences ($M = -0.34$, $SD = 0.43$ in the control condition and $M = -0.31$, $SD = 0.41$ in the experimental condition).

Explicit evaluations. The marginal means derived from the Bayesian linear model fit to the data confirmed the pattern of descriptive statistics described above. Specifically, similar levels of explicit out-group preference were observed in the control condition, $\beta = 0.46$, 95% HDI = [0.38, 0.56], and in the experimental condition, $\beta = 0.52$, 95% HDI = [0.43, 0.61]. The conditions did not differ from each other, $\beta = 0.06$, 95% HDI = [-0.07, 0.19], Bayes Factor in favor of the null hypothesis: $BF_{01} = 8.25$.

Implicit evaluations. Putting mean level differences aside, we obtained a similar pattern of results on the theoretically crucial measure of implicit evaluation. Specifically, similar levels of implicit in-group preference were observed in the control condition, $\beta = -0.64$, 95% HDI = [-0.72, -0.56], and in the experimental condition, $\beta = -0.59$, 95% HDI = [-0.67, -0.50]. The conditions did not differ from each other, $\beta = 0.05$, 95% HDI = [-0.06, 0.17], $BF_{01} = 8.13$.

Study 5B: slavery (fictitious labels, non-U.S. sample).

Descriptive statistics. Figure 6 shows the distribution of explicit and implicit evaluations, along with condition means and 95% HDIs. Participants exhibited a preference

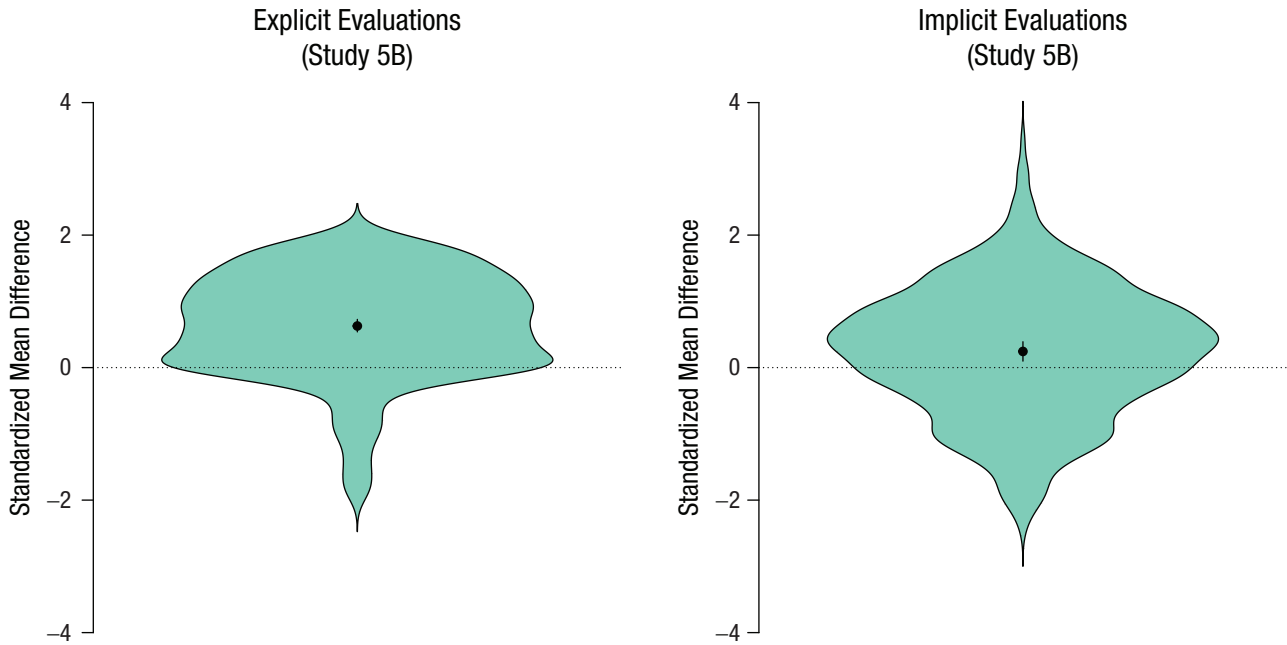


Fig. 6. Distribution of explicit and implicit evaluations by condition (Study 5B), displayed in standardized units to ensure comparability. The dashed horizontal line marks neutrality. Solid dots denote condition means. Error bars represent 95% highest density intervals. Positive scores indicate preference for the victim group over the aggressor group.

for the victim group over the aggressor group both on the measure of explicit evaluation ($M = 66.93$, $SD = 82.97$) and on the measure of implicit evaluation ($M = 0.11$, $SD = 0.45$).

Explicit evaluations. The intercept of the Bayesian mixed-effects model significantly differed from zero, $\beta_0 = 0.63$, 95% HDI = [0.54, 0.72], indicating an explicit preference for the victim group over the aggressor group.

Implicit evaluations. The intercept of the Bayesian mixed-effects model with implicit evaluations as the dependent measure also significantly differed from zero, $\beta_0 = 0.25$, 95% HDI = [0.11, 0.40], indicating an implicit preference for the victim group over the aggressor group. This result suggests that the null result obtained in Study 5A was not due to intrinsic properties of the vignette but rather the way in which White American participants in that study approached its content.

Explicit knowledge. Of 514 participants, 295 (corresponding to 57%) correctly guessed that the vignette was about slavery in the United States. Participants' self-reported knowledge of slavery in the United States was near the midpoint of the scale ($M = -1.94$, $SD = 26.13$), 95% HDI = [-4.15, 0.44]. Neither the objective nor the subjective measure was significantly associated with implicit evaluations. Explicit positivity toward the victim group was stronger as a function of both objective, $\beta = 0.19$,

95% HDI = [0.04, 0.33], and subjective, $\beta = 0.64$, 95% HDI = [0.56, 0.71], knowledge levels.

Study 5C: slavery (Black U.S. sample).

Descriptive statistics. Figure 7 shows the distribution of explicit and implicit evaluations by condition for Study 5C, along with condition means and 95% HDIs. On the measure of explicit evaluation, participants expressed an in-group preference in both conditions. This in-group preference was stronger in the experimental condition ($M = 35.44$, $SD = 43.00$) than in the control condition ($M = 25.96$, $SD = 38.74$). On the measure of implicit evaluation, participants exhibited weak in-group preference close to neutrality overall. Like in the White American sample recruited in Study 5A, we observed no condition differences ($M = 0.06$, $SD = 0.43$ in the control condition, and $M = 0.09$, $SD = 0.40$ in the experimental condition).

Explicit evaluations. The marginal means derived from the Bayesian linear model fit to the data confirmed the pattern of descriptive statistics described above. Specifically, even stronger in-group preference was observed in the experimental condition, $\beta = 0.69$, 95% HDI = [0.60, 0.77], than in the control condition, $\beta = 0.50$, 95% HDI = [0.42, 0.59]. The two conditions significantly differed from each other, $\beta = 0.19$, 95% HDI = [0.07, 0.30].

Implicit evaluations. Unlike on the measure of explicit evaluation, similar levels of weak in-group preference

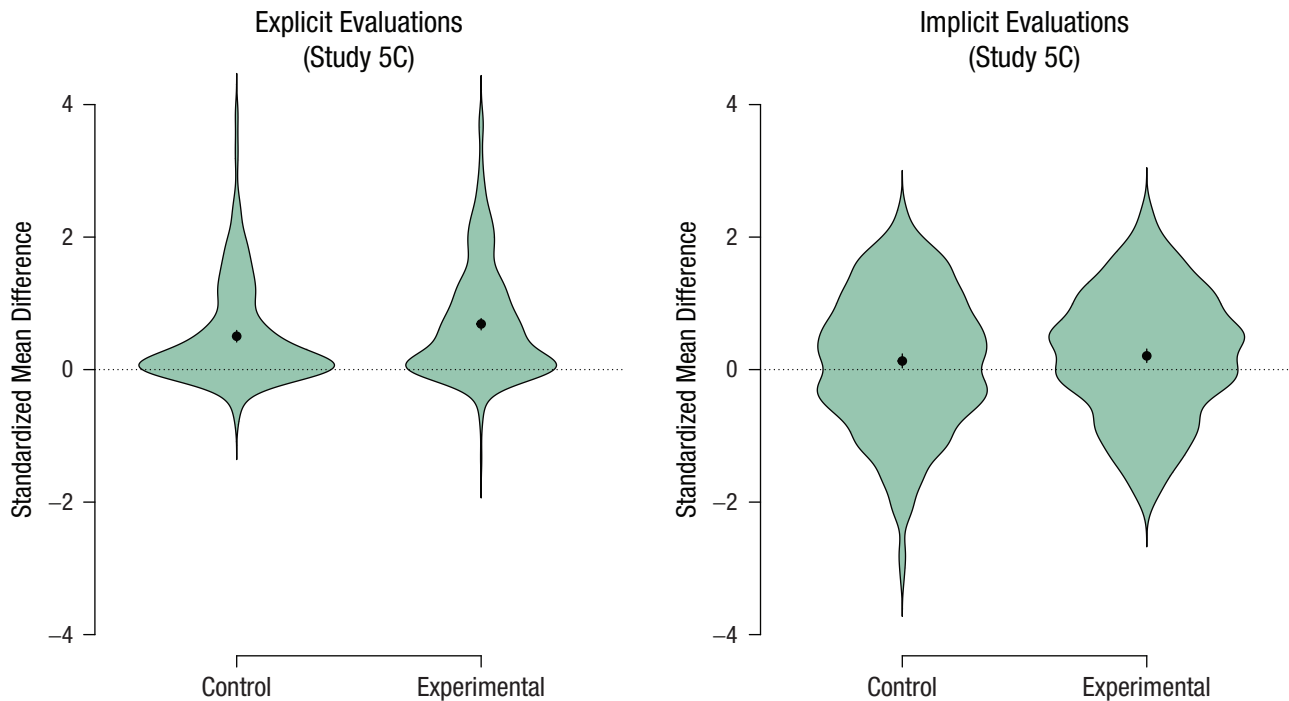


Fig. 7. Distribution of explicit and implicit evaluations by condition (Study 5C), displayed in standardized units to ensure comparability. The dashed horizontal line marks neutrality. Solid dots denote condition means. Error bars represent 95% highest density intervals. Positive scores indicate preference for Black Americans (the victim group) over White Americans (the aggressor group).

were observed both in the control condition, $\beta = 0.13$, 95% HDI = [0.02, -0.56], and in the experimental condition, $\beta = 0.21$, 95% HDI = [0.11, 0.31]. The two conditions did not differ from each other, $\beta = 0.08$, 95% HDI = [-0.06, 0.23], $BF_{01} = 7.00$.

Discussion

In Studies 3 and 4, we observed generalization of the results obtained in Studies 1 and 2 to well-known and even self-relevant targets. These results are remarkable because well-established attitudes are more difficult to change than are newly created ones (Krosnick & Petty, 1995) and because implicit evaluations are less likely to reflect relational influences in the context of well-known compared with novel targets (Kurdi et al., 2022). Moreover, in Study 4, participants' identification with the aggressor group could have thwarted learning or even produced victim derogation (Branscombe et al., 1999; Ellemers et al., 2002). Instead, we observed consistent shifts toward positivity.

Studies 5A and 5C produced a pattern of results that deviated from the remaining studies: White and Black Americans' implicit race attitudes remained unchanged following exposure to a historical narrative about slavery. Combined with a demonstration of a positive shift

in a sample of non-U.S. participants (Study 5B), these data suggest that macro-level phenomena in U.S. society can provide a parsimonious explanation of these findings. Specifically, given wide-ranging discussions of Black Americans' oppression in the context of the Black Lives Matter movement (Barrie, 2020; Reny & Newman, 2021), relevant information has likely already been incorporated into both Black and White Americans' race attitudes, thus resulting in no updating in the present studies. In contrast, bias against Native Americans is best characterized as a bias of omission (Fryberg & Eason, 2017), which can account for the strong learning effects observed in Study 4.

General Discussion

The present studies produced consistent evidence for positive shifts in implicit evaluations in response to highly negative materials describing past injustices encountered by social groups. This result is remarkable for multiple reasons. First, it suggests that prosocial (positive) responses to victims of even extreme suffering need not require effortful deliberation but instead can emerge spontaneously and unintentionally. Second, most relevant past studies—even ones involving novel targets and explicit labeling of stimulus relations—have

achieved only attenuations of implicit negativity using relational information (Kurdi et al., 2022); here, we demonstrated five instances of complete reversal into positivity. Finally, following decades of associative theorizing (Rydell & McConnell, 2006; Strack & Deutsch, 2004), this finding provides compelling support for a “new view” on which implicit evaluations reflect ubiquitous influences of high-level reasoning about causality and blameworthiness (De Houwer, 2014).

Notably, these results generalized not only across novel vignettes but even to familiar and self-relevant targets, such as the Armenians and Native Americans. This finding is of importance because convincing demonstrations of relational influences on preexisting implicit social group evaluations are virtually absent from the literature (Kurdi et al., 2022). Moreover, strong preexisting attitudes (Krosnick & Petty, 1995) and identity-related concerns (Branscombe et al., 1999; Ellemers et al., 2002) could have interfered with learning in the current studies involving real-world materials but did not.

The sole exception from this pattern was a narrative about slavery in the United States, which resulted in updating only among non-American participants but not among White or Black Americans. This finding can be parsimoniously explained by increased attention to historical and present-day anti-Black discrimination following the protests of 2020 (Abaied et al., 2022; Barrie, 2020; Reny & Newman, 2021). Given such public attention, views of Black–White race relations may have already been incorporated into race attitudes among most Black and White Americans. In line with this idea, the Black Lives Matter protests were associated with substantial attenuations of implicit in-group preference among White Americans (Sawyer & Gampa, 2018).

We note, however, that White Americans’ knowledge of Black history, and especially the history of anti-Black racism, remains limited (Bonam et al., 2019; Nelson et al., 2012). Thus, it is conceivable that updating in the positive direction may have been achieved even in the study on slavery, had we used different materials going beyond basic facts related to the enslavement of Black individuals in the United States. Such complexities notwithstanding, the results involving the slavery vignette suggest that the updating of implicit evaluations in individual minds can be modulated by macro-level processes, such as societal awareness of history. This finding, again, contradicts a view of implicit social cognition narrowly focused on stimulus associations encountered in one’s physical environment.

Taken together, the present results speak against the idea that raising awareness of past injustice creates inadvertent negativity toward oppressed groups. In fact, across seven studies, we did not find any evidence for implicit (or explicit) victim derogation, including

in the context of well-known and self-relevant social groups, either among the descendants of oppressors or the oppressed. Of course, the present studies do not eliminate the possibility of other adverse psychological consequences of encountering such materials among marginalized individuals (e.g., decreased self-esteem), thus making this area ripe for additional research.

Finally, with the finding of oppression-induced implicit positivity now firmly established, we hope that follow-up work will probe the mediators, correlates, and time course of this effect. For example, it remains to be seen whether historical narratives must possess certain linguistic or narrative features to produce implicit positivity; how best to characterize the cognitive and affective processes mediating between negative information and positive implicit responses (Lee et al., 2018); whether the learning effects produced here are instances of momentary malleability (Lai et al., 2016) or enduring change (Cone et al., 2021); and what the correlates of oppression-produced implicit positivity are, in terms of explicit evaluations (Allidina et al., 2023) and consequential societal outcomes (Payne et al., 2017). We believe that relevant inquiries will be illuminating with respect to basic processes of learning and memory subserving implicit social cognition and the question of how to confront legacies of oppression in highly unequal and stratified societies.

Transparency

Action Editor: Sylvia Perry

Editor: Patricia J. Bauer

Author Contribution(s)

Benedek Kurdi: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Writing – original draft.

Amy R. Krosch: Conceptualization; Methodology; Writing – review & editing.

Melissa J. Ferguson: Conceptualization; Methodology; Supervision; Writing – original draft.

Declaration of Conflicting Interests



Benedek Kurdi is a member of the Scientific Advisory Board of Project Implicit, a 501(c)(3) nonprofit organization and international collaborative of researchers who are interested in implicit social cognition.

Open Practices

All raw data files, analysis scripts, and materials have been made publicly available via the Open Science Framework and can be accessed at osf.io/cdftx. This article has received the badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.



ORCID iDs

Benedek Kurdi  <https://orcid.org/0000-0001-5000-0584>
 Amy R. Krosch  <https://orcid.org/0000-0002-6204-1532>

References

- Abaied, J. L., Perry, S. P., Cheaito, A., & Ramirez, V. (2022). Racial socialization messages in White parents' discussions of current events involving racism with their adolescents. *Journal of Research on Adolescence, 32*(3), 863–882. <https://doi.org/10.1111/jora.12767>
- Allidina, S., Long, E. U., Baoween, W., & Cunningham, W. A. (2023). Decoupling the conflicting evaluative meanings in automatically activated race-based associations. *Personality and Social Psychology Bulletin*. Advance online publication. <https://doi.org/10.1177/01461672231156029>
- Barrie, C. (2020). Searching racism after George Floyd. *Socius, 6*. <https://doi.org/10.1177/2378023120971507>
- Bonam, C. M., Das, V. N., Coleman, B. R., & Salter, P. (2019). Ignoring history, denying racism: Mounting evidence for the Marley hypothesis and epistemologies of ignorance. *Social Psychological and Personality Science, 10*(2), 257–265. <https://doi.org/10.1177/1948550617751583>
- Branscombe, N. R., Ellemers, N., Spears, R., & Doosje, B. (1999). The context and content of social identity threat. In N. Ellemers, R. Spears, & B. Doosje (Eds.), *Social identity: Context, commitment, content* (pp. 35–58). Wiley-Blackwell.
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Cone, J., Flaherty, K., & Ferguson, M. J. (2021). The long-term effects of new evidence on implicit impressions of other people. *Psychological Science, 32*(2), 173–188. <https://doi.org/10.1177/0956797620963559>
- Cone, J., Mann, T. C., & Ferguson, M. J. (2017). Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 56, pp. 131–199). Academic Press. <https://doi.org/10.1016/bs.aesp.2017.03.001>
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113–126. <https://doi.org/10.1037/0022-3514.44.1.113>
- De Houwer, J. (2014). A propositional model of implicit evaluation. *Social and Personality Psychology Compass, 8*(7), 342–353. <https://doi.org/10.1111/spc3.12111>
- Ellemers, N., Spears, R., & Doosje, B. (2002). Self and social identity. *Annual Review of Psychology, 53*(1), 161–186. <https://doi.org/10.1146/annurev.psych.53.100901.135228>
- Fryberg, S. A., & Eason, A. E. (2017). Making the invisible visible: Acts of commission and omission. *Current Directions in Psychological Science, 26*(6), 554–559. <https://doi.org/10.1177/0963721417720959>
- Gawronski, B., Brannon, S. M., & Ng, N. L. (2022). Debunking misinformation about a causal link between vaccines and autism: Two preregistered tests of dual-process versus single-process predictions (with conflicting results). *Social Cognition, 40*(6), 580–599. <https://doi.org/10.1521/soco.2022.40.6.580>
- Gibson, J. L. (2004). Does truth lead to reconciliation? Testing the causal assumptions of the South African truth and reconciliation process. *American Journal of Political Science, 48*(2), 201–217. <https://doi.org/10.1111/j.0092-5853.2004.00065.x>
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*(6), 1464–1480. <https://doi.org/10.1037/0022-3514.74.6.1464>
- 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), 197–216. <https://doi.org/10.1037/0022-3514.85.2.197>
- Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology, 90*(1), 1–20. <https://doi.org/10.1037/0022-3514.90.1.1>
- Hafer, C. L. (2000). Do innocent victims threaten the belief in a just world? Evidence from a modified Stroop task. *Journal of Personality and Social Psychology, 79*(2), 165–173. <https://doi.org/10.1037/0022-3514.79.2.165>
- Horwitz, S. R., & Dovidio, J. F. (2016). The rich—love them or hate them? Divergent implicit and explicit attitudes toward the wealthy. *Group Processes & Intergroup Relations, 20*(1), 3–31. <https://doi.org/10.1177/1368430215596075>
- Jordan, J. J., & Kouchaki, M. (2021). Virtuous victims. *Science Advances, 7*(42), Article eabg5902. <https://doi.org/10.1126/sciadv.abg5902>
- Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 1–24). Erlbaum.
- Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences, 14*(7), 293–300. <https://doi.org/10.1016/j.tics.2010.05.001>
- Kurdi, B., Morehouse, K. N., & Dunham, Y. (2022). How do explicit and implicit evaluations shift? A preregistered meta-analysis of the effects of co-occurrence and relational information. *Journal of Personality and Social Psychology*. Advance online publication. <https://doi.org/10.1037/pspa0000329>
- Lai, C. K., Skinner, A. L., Cooley, E., Murrar, S., Brauer, M., Devos, T., Calanchini, J., Xiao, Y. J., Pedram, C., Marshburn, C. K., Simon, S., Blanchar, J. C., Joy-Gaba, J. A., Conway, J., Redford, L., Klein, R. A., Roussos, G., Schellhaas, F. M. H., Burns, M., . . . Nosek, B. A. (2016). Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General, 145*(8), 1001–1016. <https://doi.org/10.1037/xge0000179>
- Lee, K. M., Lindquist, K. A., & Payne, B. K. (2018). Constructing bias: Conceptualization breaks the link between implicit bias and fear of Black Americans. *Emotion, 18*(6), 855–871. <https://doi.org/10.1037/emo0000347>

- Lenth, R. (2020). *emmeans: Estimated marginal means, aka least-squares means* (R package Version 1.5.0). <https://CRAN.R-project.org/package=emmeans>
- Lerner, M. J., & Miller, D. T. (1978). Just world research and the attribution process: Looking back and ahead. *Psychological Bulletin*, *85*(5), 1030–1051. <https://doi.org/10.1037/0033-2909.85.5.1030>
- Nelson, J. C., Adams, G., & Salter, P. S. (2012). The Marley hypothesis. *Psychological Science*, *24*(2), 213–218. <https://doi.org/10.1177/0956797612451466>
- Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*(4), 233–248. <https://doi.org/10.1080/1047840x.2017.1335568>
- R Core Team. (2022). *R: A language and environment for statistical computing* (Version 4.2.1) [Computer software]. R Foundation for Statistical Computing. <http://www.R-project.org>
- Reny, T. T., & Newman, B. J. (2021). The opinion-mobilizing effect of social protest against police violence: Evidence from the 2020 George Floyd protests. *American Political Science Review*, *115*(4), 1499–1507. <https://doi.org/10.1017/s0003055421000460>
- Rotella, K. N., & Richeson, J. A. (2013). Motivated to “forget”: The effects of in-group wrongdoing on memory and collective guilt. *Social Psychological and Personality Science*, *4*(6), 730–737. <https://doi.org/10.1177/1948550613482986>
- Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology*, *91*(6), 995–1008. <https://doi.org/10.1037/0022-3514.91.6.995>
- Sawyer, J., & Gampa, A. (2018). Implicit and explicit racial attitudes changed during Black Lives Matter. *Personality and Social Psychology Bulletin*, *44*(7), 1039–1059. <https://doi.org/10.1177/0146167218757454>
- Southern Poverty Law Center. (2018). *Teaching hard history*. <https://www.splcenter.org/20180131/teaching-hard-history>
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*(3), 220–247. https://doi.org/10.1207/s15327957pspr0803_1
- Uhlmann, E. L., Brescoll, V. L., & Paluck, E. L. (2006). Are members of low status groups perceived as bad, or badly off? Egalitarian negative associations and automatic prejudice. *Journal of Experimental Social Psychology*, *42*(4), 491–499. <https://doi.org/10.1016/j.jesp.2004.10.003>
- Zanon, R., De Houwer, J., Gast, A., & Smith, C. T. (2014). When does relational information influence evaluative conditioning? *Quarterly Journal of Experimental Psychology*, *67*(11), 2105–2122. <https://doi.org/10.1080/17470218.2014.907324>