Journal of Experimental Psychology: General

A Comparative Investigation of Interventions to Reduce Anti-Fat Prejudice Across Five Implicit Measures

Calvin K. Lai and Joel M. Le Forestier

Online First Publication, February 3, 2025. https://dx.doi.org/10.1037/xge0001719

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

https://doi.org/10.1037/xge0001719

A Comparative Investigation of Interventions to Reduce Anti-Fat Prejudice Across Five Implicit Measures

Calvin K. Lai[1] and Joel M. Le Forestier[2]
[1] Department of Psychology, Rutgers University
[2] Department of Psychology, University of Pittsburgh

The severity and pervasiveness of anti-fat prejudice and discrimination have led to calls for interventions to address them. However, intervention studies to combat anti-fat prejudice have often been stymied by ineffective approaches, small sample sizes, and the lack of standardization in measurement. To that end, we conducted two mega-experiments totaling 27,726 participants and 50 conditions where we tested five intervention approaches to reduce implicit anti-fat prejudice across five implicit measures. We found that interventions were most effective at reducing implicit weight biases when they instructed people to practice an explicit rule linking fat people with good things and thin people with bad things. Interventions that were more indirect or relied on associative learning tended to be ineffective. We also found that change in implicit bias on one implicit measure often generalized to other implicit measures. However, the Evaluative Priming Task and single-target measures of implicit bias like the Single-Target Implicit Association Test were much less sensitive to change. These findings illuminate promising approaches to combating implicit anti-fat prejudice and advance understanding of how implicit bias change generalizes across measures.

Public Significance Statement

Anti-fat stigma is pernicious, common, and often seen as socially acceptable. As a result, fat people contend with workplace discrimination, marginalization in health care, and victimization by close others that degrade their well-being. Anti-fat stigma touches the lives of countless people, but the best ways to address it are unknown, as interventions to reduce anti-fat prejudice have often failed to work reliably. We conducted two large-scale experiments systematically testing five promising interventions to reduce implicit anti-fat prejudice. The best interventions encouraged the practice of an explicit rule linking fat people with good things and thin people with bad things. Ineffective interventions were less direct in targeting prejudice, such as an intervention that reminded people of the general importance of being egalitarian. These findings suggest that effective interventions to combat implicit anti-fat prejudice will involve messaging that directly connects fatness with positivity and that relying on indirect messaging will be insufficient.

Keywords: weight prejudice, interventions, social cognition, implicit measures, intergroup relations

Supplemental materials: https://doi.org/10.1037/xge0001719.supp

Prejudice and discrimination against fat people are pernicious and deeply entrenched. Fat people are often stereotyped negatively as lazy, lacking in willpower, and unintelligent (Crandall, 1994). As a result, fat people struggle with family members and peers who mock them for their weight (Puhl et al., 2008), face hiring discrimination (Campos-Vazquez & Gonzalez, 2020; Rooth, 2009), and contend with medical professionals who reinforce weight stigma in their treatment (Puhl & Heuer, 2010). The experience of discrimination is

Tara Mandalaywala served as action editor.

Calvin K. Lai https://orcid.org/0000-0003-2437-9783

Preregistration, data, and materials for this article are publicly available on the Open Science Framework at https://osf.io/nw5hv/. The ideas and data of the article were presented at the Person Memory Interest Group in 2019 and the Society of Personality and Social Psychology in 2022. Calvin K. Lai is a consultant and member of the Scientific Advisory Board with Project Implicit, a nonprofit organization and international collaborative of researchers who are interested in implicit social cognition.

Conceptualization: Calvin K. Lai, Joel M. Le Forestier; data curation: Calvin K. Lai, Joel M. Le Forestier; formal analysis: Calvin K. Lai, Joel M. Le Forestier; funding acquisition: Calvin K. Lai; investigation: Calvin K. Lai, Joel M. Le Forestier; methodology: Calvin K. Lai, Joel M. Le Forestier; project administration: Calvin K. Lai, Joel M. Le Forestier; resources: Calvin K. Lai; software: Joel M. Le Forestier, Calvin K. Lai; supervision: Calvin K. Lai; validation: Calvin K. Lai, Joel M. Le Forestier; visualization: Calvin K. Lai, Joel M. Le Forestier; writing–original draft: Calvin K. Lai; writing–review and editing: Calvin K. Lai, Joel M. Le Forestier.

Calvin K. Lai played a lead role in conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, visualization, writing–original draft, and writing–review and editing and a supporting role in software. Joel M. Le Forestier played a lead role in software and a supporting role in conceptualization, data curation, formal analysis, investigation, methodology, project administration, validation, visualization, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to Calvin K. Lai, Department of Psychology, Rutgers University, 53 Avenue East, Piscataway, NJ, 08854, United States. Email: calvin.lai@rutgers.edu

made worse by the widespread acceptance of anti-fat discrimination (Puhl & Brownell, 2001). Often, people attempt to justify blatant expressions of weight stigma as important for motivating others to lose weight (Hunger et al., 2020; Logel et al., 2015). Ironically, the evidence finds the opposite: Weight stigmatization denigrates the physical and mental well-being of fat people rather than ameliorating them (Hunger et al., 2020; Puhl & Heuer, 2009).

The extent of anti-fat prejudice is so severe that even the adjective "fat" is often seen as pejorative (Brochu & Esses, 2011). In our article, we use the adjective "fat" to align with efforts to reclaim the word "fat" as a neutral or positive term (Galinsky et al., 2003; Saguy & Ward, 2011; see Footnote 1 for a detailed definition of our other key terms).[1] We avoid the use of alternative terms like "overweight" or "obese" due to their pathologizing medical origins.

Underscoring the severity of anti-fat prejudice, a study of 4.4 million people from 2007 to 2016 across six intergroup domains (i.e., sexual orientation, race, skin tone, age, disability, weight) found that most forms of implicit prejudice were declining or remaining stable except for one: weight (Charlesworth & Banaji, 2019). Implicit anti-fat prejudice was the only form of implicit prejudice increasing over time. Furthermore, anti-fat prejudice has also been found in all 71 nations studied in a large-scale cross-cultural study (Marini et al., 2013).

Given the prominence of anti-fat prejudice, researchers and practitioners alike have sought to develop interventions that can reliably combat it. However, reviews of weight bias intervention studies have shown limited evidence for effective interventions to reduce anti-fat prejudice (Alberga et al., 2016; Daníelsdóttir et al., 2010; Lee et al., 2014; Talumaa et al., 2022). A review by Daníelsdóttir et al. (2010) concluded that interventions often changed beliefs about weight without creating corresponding decreases in anti-fat prejudice. Another review, a meta-analysis, found only 30 relevant studies (Lee et al., 2014). Of these, 19 studies used experimental designs and had a median of 39 participants per condition. The remaining 11 studies used quasi-experimental designs and were unable to make strong causal inferences (e.g., comparing two different undergraduate courses).

The dearth of reliable evidence on intervention efficacy is exacerbated by the lack of standardization in prejudice measures. Without standardization, it is difficult to compare which interventions are working and which are not. Daníelsdóttir et al. (2010) identified 21 weight prejudice intervention studies that assessed outcomes on 22 distinct weight prejudice measures. Sixteen of these 22 measures were only used a single time. This lack of standardization has persisted over time, as a recent review of weight stigma reduction studies in health care finds that 26 out of 54 outcomes in the literature employed measures that were only used once (Talumaa et al., 2022).

Across the reviews, the most common measure of weight prejudice was a measure of implicit weight prejudice: the Weight Implicit Association Test (IAT; Nosek et al., 2007). Despite being a common target for change, six out of seven interventions from published experiments failed to consistently change Weight IAT scores (Gapinski et al., 2006; Matharu et al., 2014; O'Brien et al., 2010; Swift et al., 2013; Teachman et al., 2003). The interventions that failed to change Weight IAT scores evoked empathy toward fat people (Gapinski et al., 2006; Matharu et al., 2014; Teachman et al., 2003), challenged stereotypes by giving people experiences with counter-stereotypical fat people (Gapinski et al., 2006; Swift et al., 2013), and corrected misconceptions about weight as controllable and exclusively about individual behaviors (Matharu et al., 2014; Teachman et al., 2003). The only effective intervention was an intensive intervention conducted over 3 weeks that also corrected misconceptions about weight as controllable and highlighted the role of uncontrollable factors like genetics (O'Brien et al., 2010).

The lack of consistent evidence for reductions in implicit weight prejudice was accompanied by small sample sizes; the average sample size for experimental conditions in these studies was 39 people. This meant that the average study only had 80% power to detect effects of $d = 0.64$. By contrast, the average effect size in experimental research on implicit bias change is about $d = 0.25$ (Forscher et al., 2019), which would require an average sample size of 253 people per condition to detect. This suggests that studies in research on implicit weight prejudice reduction have been grossly underpowered.

The amount of evidence about implicit weight prejudice reduction also pales in comparison to the amount of research on other domains such as implicit racial prejudice reduction. In a meta-analysis of 492 experiments to change implicit bias, Black/White racial biases comprised 31% of all experimental conditions studied, whereas weight biases only comprised 1% (Forscher et al., 2019). This is notable given that the unique structure of race relations and racial attitudes may not generalize to weight attitudes. The majority of Americans do not identify as Black and contact with Black Americans is relatively rare given their status as a numerical minority and the extent of racial segregation in the United States. By comparison, fat people comprise a majority of the U.S. population and are relatively more widely dispersed. Self-reported prejudice against fat people is also generally higher than self-reported prejudice against Black people (Nosek et al., 2007), and blatant expressions of prejudice against fat people are generally seen as more acceptable than expression of prejudice against Black people (Crandall et al., 2002). Such differences in the acceptability of prejudice are also shown in the relative lack of ingroup biases among fat people (Crandall, 1994). Anti-fat and anti-Black prejudice also arise from distinct forms of negative evaluation. Whereas anti-fat prejudice is often tied to disgust (Vartanian, 2010; Vartanian et al., 2013), anti-Black prejudice is more commonly tied to physical threat and aggression (March et al., 2021; Sim et al., 2022). In sum, studying implicit attitudes about fat people will test the generalizability of findings about anti-Black implicit attitudes to a group with whom more people regularly interact, are more likely to express blatant prejudice against, and are more likely to identify with but less likely to show ingroup biases for and for which prejudice is believed to come from different origins (e.g., disgust).

[1] In this article, we define "implicit measures" as measures that indirectly assess mental content through performance in an ostensibly unrelated task (e.g., Implicit Association Test; Lai & Wilson, 2021). "Implicit bias" and "implicit preference" are used interchangeably to refer to behavioral response tendencies on implicit measures. Finally, we define "attitudes" as an umbrella construct encompassing valenced (i.e., positive/negative) evaluations and "prejudices" as the subset of attitudes about social groups.

When we use these terms, we do take a theory-uncommitted view and do not make strong assumptions about the nature of the underlying structures (Greenwald et al., 2005). The structures may reflect enduring representations (e.g., Wilson et al., 2000), assessments constructed on-the-spot (Schwarz, 2007), or propositions (De Houwer, 2014).

In the present research, we conducted two large-scale experiments to reduce implicit weight prejudice that address each of the limitations of research on weight prejudice reduction. First, we tested multiple interventions in experiments while controlling for the sample, setting, measure, and procedural details that are unrelated to the intervention itself. Doing so allowed us to make strong causal inferences about what interventions are most effective at changing implicit weight biases. Second, we tested the generalizability of implicit weight bias reduction by examining efficacy across five of the most commonly used measures in research on implicit intergroup biases (Lai & Wilson, 2021): the IAT, Brief IAT (BIAT), Affect Misattribution Procedure (AMP), Evaluative Priming Task (EPT), and Single-Target IAT (ST-IAT). Third, our ability to make strong causal inferences was strengthened by sample sizes that were many times larger than data that had been collected over the past 2 decades. Our combined analytic sample across both experiments was 27,114, which is about 37 times as much data as had been previously published in prior anti-fat prejudice reduction studies combined (total $N = 736$). Finally, recognizing the limited efficacy of past interventions to reduce implicit weight prejudice, we tested interventions that were previously effective at reducing implicit biases in other intergroup domains (e.g., Kurdi & Banaji, 2017; Lai et al., 2014).

Interventions to Reduce Implicit Weight Prejudice

The theoretical and empirical foundations of each tested intervention are described below. In selecting these interventions, we sampled theoretical approaches that have primarily been tested on novel social groups (evaluative statement, passive evaluative conditioning), on real-world social groups (emotional narrative, egalitarian goals), and a mixture of both (active evaluative conditioning).

Emotional Narrative (Studies 1 and 2; Adapted From Lai et al., 2014; Marini et al., 2012)

Experiences with group members who defy one's preexisting prejudicial attitudes can be a powerful lever for changing implicit attitudes (Lai & Banaji, 2020; Lai et al., 2013). These counter-attitudinal experiences may be especially powerful when the experiences are emotionally vivid and self-relevant (Calanchini et al., 2021). To that end, participants in this intervention read a second-person story where they experience the horror of being attacked by a villainous thin man. However, they are ultimately rescued by a heroic fat man who fights off the thin man and saves the day. This intense emotional experience was theorized to activate mental content that was contrary to pro-thin/anti-fat attitudes (Lai et al., 2014; Marini et al., 2012).

Egalitarian Goals (Study 1; Adapted From Moskowitz & Li, 2011)

Inducing the goal to egalitarian can activate goal shielding that inhibits implicit biases. Rather than changing the underlying attitudes, efforts to induce egalitarian goals were theorized to prevent the activation of prejudicial attitudes to begin with (Moskowitz & Li, 2011). To that end, participants in this intervention wrote about a time they failed to live up to their egalitarian ideals in relation to fat people.

Active Evaluative Conditioning (Study 1; Adapted From Schmidt & de Houwer, 2012)

Evaluative conditioning tasks repeatedly expose people to pairings of a stimulus with a positive or negative stimulus. These tasks are often theorized to change the attitudes toward that stimulus through mental tracking of co-occurrence (Hofmann et al., 2010; Kurdi et al., 2023). In this intervention, participants took a special variant of evaluative conditioning that required participants to actively respond to stimulus pairings reinforcing that fat = good and thin = bad by categorizing what they saw as being about fat or thin people. A version of this active evaluative conditioning task has generated large changes in attitudes (Gast & Rothermund, 2010; Schmidt & de Houwer, 2012; cf. Gast et al., 2020). We theorized that this active conditioning approach would complement learning through co-occurrence with learning through the practice of an explicit rule linking fat people with good things and thin people with bad things.

Passive Evaluative Conditioning (Study 2; Adapted From Kurdi & Banaji, 2017)

In contrast to the active evaluative conditioning intervention where participants were asked to actively respond to stimulus pairings, participants in this intervention were passively exposed to stimulus pairings reinforcing that fat = good and thin = bad. As with active evaluative conditioning, exposure to counterattitudinal pairings was theorized to shift implicit biases. This intervention approach had often generated relatively small changes in implicit attitudes but is more commonly used in research on evaluative conditioning (Kurdi et al., 2023). This intervention also served as a more conservative test of attitude change through mental tracking of co-occurrence, as it did not have the active responding component of the active evaluative conditioning intervention.

Evaluative Statement (Study 2; Adapted From Kurdi & Banaji, 2017)

Evidence over the past decade has demonstrated that implicit measures are often more sensitive to rule-based information about the relationships between concepts than mere associative co-occurrence of the kind studied in evaluative conditioning interventions (Cone et al., 2017; Kurdi & Dunham, 2020; Kurdi et al., 2023; Kurdi & Mandelbaum, 2023). Implicit attitudes can be incredibly responsive to reasoning, logic, and rule-based instructions. To that end, participants were simply given an instruction that they will see stimulus pairings reinforcing that fat = good and thin = bad.

The Measurement of Change in Implicit Biases

We examined the differential sensitivity of implicit measures to change by testing the effect of each intervention on five implicit measures. Doing so informed theory about the measurement and interpretation of implicit measures while also informing our understanding of the robustness of weight bias reduction interventions across operationalizations of implicit bias.

Our motivation to study efficacy across implicit measures is informed by a meta-analysis of change in implicit measures that found a majority (65%) of experiments to change implicit bias relied

on the IAT (Forscher et al., 2019). However, it is well understood that using measures beyond the IAT in bias reduction research can tease apart psychological processes that are not possible with just the IAT (Gawronski, 2023; Lai et al., 2013). For example, the Weight IAT is designed to measure relative preferences between fat and thin people and relies on categorization to semantic categories. It also assesses implicit attitudes through reaction time, which may capture the controllability and speed of mental processing more than other automatic processes (e.g., unconscious processing). Other implicit measures can measure absolute attitudes about a single-target group (e.g., ST-IAT; Bluemke & Friese, 2008), use priming approaches that do not rely on group-based categorization (e.g., EPT; Fazio et al., 1995), or assess automatic processes through evaluative judgments that may better capture unconscious attitudes (e.g., AMP; Payne et al., 2005).

Due to the theoretical differences in what implicit measures assess, measures may differ in their sensitivity to change. To date, there has been little direct comparative work examining whether implicit measures are more sensitive to some sources of change than others (cf. Bar-Anan & Nosek, 2016; Deutsch & Gawronski, 2009; Gawronski & Bodenhausen, 2005). That may matter, as interventions vary in whether they appeal to emotions or beliefs, and some measures may be more sensitive to emotional or cognitive-based sources of change. Some early theorizing suggested that the affect misattribution procedure may be more sensitive to affective processes (Payne et al., 2005; cf. Blaison et al., 2012), but research has not directly compared AMP's sensitivity with affective sources of change compared with other measures.

Interventions to change intergroup biases can also vary in whether they primarily present associative information or propositional information about the relations between concepts (Kurdi & Banaji, 2017). Many earlier studies have shown that the IAT is especially sensitive to associative information in the environment (for a review, see Olson et al., 2009). Recent meta-analytic evidence has also found that studies using the AMP find larger learning effects from propositional information than studies using the EPT or IAT-based measures (Kurdi et al., 2023). More direct research is needed to examine which implicit measures are more sensitive to propositional versus associative information.

Additionally, measures may vary in their susceptibility to intentional goals to respond without bias. The AMP has appeared to be the most susceptible to spontaneous goals to respond without bias (Bar-Anan & Nosek, 2012, 2014; Gawronski & Ye, 2014; T. C. Mann et al., 2019). The IAT and EPT are also susceptible to intentional goals to respond without bias, but doing so usually requires explicit instruction on how to do so (Banse et al., 2001; Czellar, 2006; Degner, 2009; Fiedler & Bluemke, 2005; Kim, 2003; Teige-Mocigemba & Klauer, 2008). However, without studies that directly compare measures, it is a challenge to make relative comparisons in measures' susceptibility to intentional goals.

A final possibility is that implicit measures operate similarly in detecting psychological change. Oftentimes, implicit measures load onto the same construct in latent variable modeling (Bar-Anan & Vianello, 2018; Cunningham et al., 2001; Ranganath et al., 2008). To the extent that this is the case, differences in sensitivity to change could be attributable to measurement-related differences in internal consistency. According to several large-scale studies comparing implicit measures (Bar-Anan & Nosek, 2014; Greenwald & Lai, 2020; Lai & Wilson, 2021), IAT-based measures had the highest internal consistency, and the EPT had one of the lowest. That could mean that IAT-based measures are most sensitive to psychological change, whereas noisier measures like the EPT are the least sensitive.

Study 1

We tested the efficacy of three interventions that represented major approaches to changing implicit attitudes: an emotional narrative that gave people emotionally evocative experiences with an evil thin man and a heroic fat man, an active evaluative conditioning procedure that exposed participants to stimulus pairings that reinforced that fat = good and thin = bad, and an egalitarian goals intervention that led participants to care more strongly about being fair to fat people. These interventions had previously been effective at changing implicit racial biases (Lai et al., 2014; Marini et al., 2012; Moskowitz & Li, 2011). We also measured participants' self-reports about theoretically relevant mechanisms for each of these interventions.

These interventions were tested on five implicit measures. Four of the measures were oriented toward measuring relative preferences between fat and thin people and were the primary focus of our analyses. A fifth measure (the ST-IAT) exclusively measured biases about a single social group and was used for secondary analyses to examine whether interventions were effective at changing single-target biases along with two other measures.

Method

Participants

Our final analytic sample included 14,380 participants from the research platform Project Implicit (https://implicit.harvard.edu). The final sample was 69% female and 71% male. The mean age was 33.6, and 68% of participants identified as White, 8% as Black, 6% as Asian, 12% as Hispanic/Latino, 0.5% as Native Hawaiian or Pacific Islander, 0.4% as American Indian/Alaska Native, 4% as multiracial, and 1% as other or unknown. For the highest level of education, 2% had less than a high school degree, 33% had a high school degree, 39% had an undergraduate degree, and 27% had a postgraduate degree. Politically, 53% identified as liberal, 28% identified as moderate, and 19% identified as conservative.

We originally planned to collect an average of 749 participants per condition, totaling 14,980 participants across the 20 conditions. That would mean 80% power to detect mean differences of d = 0.15 between an intervention condition and the control condition (assuming 7% data excluded). Twenty-two thousand thirty-eight American residents or citizens consented to take the study, of which 14,992 (68.0%) completed the study. We then excluded 612 participants (4.1%) who violated implicit measure exclusion criteria for our final sample of 14,380. This consisted of 24/2956 (0.8%) participants who took the IAT, 56/2903 (1.9%) participants who took the BIAT, 24/3049 (0.8%) participants who took the EPT, 485/2959 (16.4%) participants who took the AMP, and 23/3036 (0.7%) participants who took the ST-IAT. We followed the exclusion criteria used by Bar-Anan and Nosek's (2014) study of the psychometrics of seven implicit measures. For detailed information

on exclusion criteria and how demographic information was collected, see additional online material at https://osf.io/tdr2g.

Procedure

Participants were randomly assigned to one of 20 conditions in a 4 (intervention: emotional narrative, active evaluative conditioning, egalitarian goals, baseline control) × 5 (implicit measure: IAT, BIAT, AMP, EPT, ST-IAT) design. Participants in the baseline control condition did not complete any task prior to completing an implicit measure. After taking an intervention or baseline control, participants completed the implicit measure they were randomly assigned to. Finally, they completed self-report measures of weight attitudes, potential intervention-specific mechanisms, and several additional exploratory measures (i.e., perceived control over body weight and indices of subjective and objective weight status).

Emotional Narrative (Adapted From Lai et al., 2014; Marini et al., 2012). Participants read an intense emotional story where they imagined walking down a street late at night after drinking at a bar. Suddenly, a thin man assaults them, throws them into the trunk of his car, and drives away. After some time, the thin man opens the trunk and assaults the protagonist again. A heroic fat man notices this assault and knocks out the villain, saving the day.

Active Evaluative Conditioning (Adapted From Schmidt & de Houwer, 2012). Participants actively responded to stimulus pairings reinforcing that fat = good and thin = bad. They first received instructions about the learning task and were told to categorize images of fat and thin people as quickly as possible. They were then presented conditioning trials across three blocks of 30 trials each. In each trial, participants saw a fat silhouette paired with a positive word or a thin silhouette paired with a negative word. Then, they had to actively categorize the silhouette as either "fat" or "thin." This act of categorization was theorized to give participants practice with the rule that fat = good and thin = bad. There was an intertrial interval of 1,000 ms. If the participant responded incorrectly, they would see a red "X" alongside the image and word for 300 ms before the intertrial interval. The silhouettes and valenced words were the same stimuli as the ones used in the implicit measures.

Egalitarian Goals (Adapted From Moskowitz & Li, 2011). Participants were given a list of 12 goals and traits and asked to rate how important they found each of those traits (e.g., egalitarianism, respect for tradition, pleasure, intelligence). Each participant was then told they would focus on one of the goals on the last page. All of them were assigned to focus on "egalitarianism." They then read a definition of egalitarianism and wrote about a time they failed to live up to the ideal of egalitarianism relating to fat people.

Implicit Measures

To standardize measurement, we adopted near-identical stimuli, category labels, and scoring procedures for all implicit measures. The measures were all taken from Bar-Anan and Nosek's (2014) comparative investigation of the psychometrics of implicit measures, with slight deviations from their procedure described below. Full details of the procedures and scoring for these implicit tasks are at https://osf.io/nw5hv/.

Category Labels and Stimuli. The category labels for the three IAT-based measures were "fat people," "thin people," "good," and "bad." The labels for the AMP and EPT were "good" and "bad." Stimuli of fat and thin people for all measures were silhouettes of fat and thin people taken from the Weight IAT at Project Implicit (Xu et al., 2014). These silhouettes were designed to be visually prototypical of fat and thin people and had been previously validated in large-scale studies of hundreds of thousands of people (Charlesworth & Banaji, 2019; Marini et al., 2013; Nosek et al., 2007). Attribute stimuli were the words "joy, glorious, wonderful, love, happy, laughter, pleasure, and peace" for "good" and the words "terrible, nasty, evil, hurt, horrible, failure, awful, and agony" for "bad."

Implicit Association Test (Greenwald et al., 1998). In two initial practice blocks, participants saw images of fat and thin people on the screen and categorized then as "fat people" or "thin people" (Block 1; 20 trials) and saw valenced words and categorized them as "good" or "bad" (Block 2; 20 trials) by sorting them to the left or right side of the screen. In critical Blocks 3 (20 trials) and 4 (40 trials), participants categorized fat people and good words with one key and thin people and bad words with the other key. Block 5 (40 trials) was another practice block that just involved categorizing fat people and thin people, but with the sides switched. Finally, in critical Blocks 6 (20 trials) and 7 (40 trials), participants categorized thin people and good words with one key and fat people and bad words with the other key. The order of the critical blocks was randomized to control for potential order effects. We deviated from Bar-Anan and Nosek's (2014) procedure to use 28 trials on practice Block 5 instead of 40 trials to match the current standard practice at the Project Implicit research platform.

Brief IAT (Sriram & Greenwald, 2009). BIAT's stimuli and categorization decision was similar to the IAT, but the framing of the task and block structure was different. Framing-wise, participants were told they would be playing an "in-our-out game." They would see a focal social category ("fat people" or "thin people") and the focal attribute category "good" in the center of the screen instead of four categories in the IAT. Participants were then instructed to hit the right key if the word/image presented on the screen belonged in one of the focal categories (in) or the left key if the word/image presented on the screen did not belong in one of the focal categories (out). The nonfocal stimuli were always bad words and the group not mentioned (i.e., thin people if the focal social category was "fat people" and fat people if the focal social category was "thin people").

Participants completed four critical blocks at 20 trials each. In the first two critical blocks (2 and 3), participants categorized the focal categories as "fat people" and "good." In the second two critical blocks (4 and 5), the focal categories were "thin people" and "good." The order of the critical blocks was randomized to control for potential order effects.

Affect Misattribution Procedure (AMP; Payne et al., 2005). Participants saw a prime image on each trial (i.e., a silhouette of a fat person or a thin person), followed by an abstract painting and a mask. Participants were then instructed to judge the abstract painting as either more pleasant or unpleasant than the average abstract painting. The measure contained three practice trials and 100 critical trials. We deviated from Bar-Anan and Nosek's (2014) procedure to follow newer best practices to prevent intentional responding (T. C. Mann et al., 2019). These

best practices meant using abstract paintings as the stimuli-to-be-judged rather than Chinese characters and adding an extra instruction page that emphasized the importance of evaluating the paintings rather than the primes.[2] We also used "good" and "bad" as the category labels instead of the more conventional "pleasant" and "unpleasant" to match the labels in the other implicit measures.

Evaluative Priming Task (Fazio et al., 1995). In an initial practice block containing 28 trials, participants categorized words as "good" or "bad." Then, they completed three critical blocks at 60 trials each where they continued to sort words as "good" or "bad" but saw a prime image before each word. The prime images were images of fat and thin people.

Single-Target IAT (Bluemke & Friese, 2008). The ST-IAT was similar to the IAT but only assessed fat people. After an initial practice block to familiarize participants with the "good" and "bad" categories, participants completed four critical blocks of 48 trials each. In the first two critical blocks (2 and 3), participants categorized fat people and good words with one key and bad words with the other key. In the second two critical blocks (4 and 5), participants categorized fat people and bad words with one key and good words with the other key. The order of the critical blocks was randomized to control for potential order effects. Unlike in Bar-Anan and Nosek's (2014) study, we did not also conduct an additional ST-IAT assessing another group (thin people) due to timing constraints.

Scoring for Relative Preferences. Our primary analyses focused on a relative preference score between thin and fat people, with positive scores indicating preferences for thin people over fat people. We applied trial- and session-based exclusion criteria for all measures based on Bar-Anan and Nosek's (2014; see https://osf.io/nw5hv/) study.

Following Bar-Anan and Nosek (2014), we relied on a general scoring procedure called *D* for the IAT and BIAT. *D* has been validated as a useful approach for measuring contrasting conditions that use response latency as a dependent variable, with validation work on the IAT and BIAT suggesting it helps optimize correlations with explicit measures and mitigates contamination by individual differences in reaction time (Greenwald et al., 2003; Nosek et al., 2014). *D* can be summarized with the following (Equation 1; Greenwald et al., 2003):

$$D = (M_{\text{Fat people + Good/Thin people + Bad}} - M_{\text{Thin people + Good/Fat people + Bad}})/SD. \quad (1)$$

D standardizes average differences in response latency by dividing the difference between the average latency of one set of trials ($M_{\text{Fat people + Good/Thin people + Bad}}$) and the average latency of the *other* set of trials ($M_{\text{Thin people + Good/Fat people + Bad}}$) by the standard deviation of response latencies across both sets of trials. For analysis, we relied on an average *D* score that was computed as an average of two *D* scores from the two halves of the IAT and BIAT.

Following Bar-Anan and Nosek's (2014) EPT scoring procedure, we computed average log-transformed response latencies for each of the four types of trials (i.e., fat people + good, fat people + bad, thin people + good, thin people + bad). We then computed adapted *D* scores for each social category (e.g., ($M_{\text{Fat people + Good}} - M_{\text{Fat people + Bad}}$)/*SD*). We computed difference scores for each of the three blocks by taking the difference between the two social category scores. Finally, we computed an average of those three differences in scores of the EPT for our primary analyses.

As the AMP relied on relative differences in responses rather than response latency, we assessed relative preference by subtracting the rate of "pleasant" responses after primes of fat people from the rate of "pleasant" responses after primes of thin people.

Scoring for Single-Target Biases. We computed indices of single-target biases for three measures: the AMP, EPT, and ST-IAT. Positivity toward fat people on the AMP was assessed by the proportion of "pleasant" responses after primes of fat people, and positivity toward thin people was assessed by the proportion of "pleasant" responses after primes of thin people. Single-target biases for the EPT were assessed similarly to our measure of relative preference, but we took the average of the three adapted *D* scores for each social category instead of the average of the difference scores. The ST-IAT only assessed single-target biases about fat people. The scoring procedure was similar to the IAT and BIAT but yielded a single-target bias score due to the block structure of the ST-IAT. Each ST-IAT *D* score was computed as $D = (M_{\text{Fat people + Good}} - M_{\text{Fat people + Bad}})/SD$.

Explicit Measures

In addition to implicit measures, we also assessed self-reported weight attitudes, intervention-specific moderators for each of the three intervention conditions, and several additional exploratory measures.

Explicit Weight Attitudes. Participants completed three self-report items measuring weight attitudes. One assessed relative preference for thin people over fat people on a 7-point Likert scale ranging from −3 (*I strongly prefer fat people to thin people*) to +3 (*I strongly prefer thin people to fat people*). The others were feeling thermometers for thin people and fat people measured using a 7-point scale ranging from −3 (*very cold*) to +3 (*very warm*). For analyses, a difference score was computed between the two feeling thermometers and averaged with the weight preference measure after standardizing each ($SD = 1$) while retaining a rational zero point of no preference between thin people and fat people. Higher positive scores indicated a greater explicit preference for thin people over fat people.

Intervention-Specific Mechanisms. Participants in the emotional narrative condition were asked questions about their emotional reactions to the story. Specifically, they were asked how emotionally affected, distressed, upset, scared, disgusted, and relieved by the story on a scale from 1 (*not at all [mental state]*) to 5 (*extremely [mental state]*). We analyzed these data with all of the questions averaged together ($\alpha_{\text{Study 1}} = .87$, $\alpha_{\text{Study 2}} = .88$).

Participants in the active evaluative conditioning condition were asked questions to test their memory for the contingencies between stimuli. They were shown each of the eight images in the intervention and asked to select what type of words they were paired with:

[2] We had learned of these best practices from T. C. Mann et al.'s (2019) article before it had left peer review, so we sought to test them ourselves before using them in Study 1. We conducted a small preregistered pilot experiment (total $N = 307$; 77 per condition) examining the impact of the extra instruction page for the AMP and EPT. The survey materials, data, and a report on the findings are available at https://osf.io/nw5hv/ and the preregistration is available at https://osf.io/cthq6/.

"positive words," "negative words," "neutral words," or "unsure." Due to a programming error, participants who took the EPT did not receive these questions.

Participants in the egalitarian goals condition were asked two questions about their motivation to be egalitarian: "How important is it to you to be egalitarian towards fat people right now?" and "Right now, how important is it to you that people are treated equally regardless of their body weight?" on scale from 1 (*not important at all*) to 5 (*extremely important*). For analysis, answers to both questions were averaged together ($\alpha = .79$).

Additional Exploratory Measures. In both studies, we also assessed perceived control over body weight as an exploratory outcome and three indices of body weight status (subjective weight status, perceived similarity to fat and thin people, objective weight status) as potential moderators of intervention effects. We found that the interventions did not affect perceived control over body weight and that indices of body weight status did not moderate intervention effects in both studies. A detailed description of these measures and analyses is available in the Supplemental Materials and additional online material at https://osf.io/tdr2g.

Transparency and Openness

We report all data exclusions, manipulations, measures, and how we determined our sample sizes. Decisions to stop collecting data did not depend on the results we obtained. All data, analysis scripts, survey materials, and additional online material have been made publicly available at the Open Science Framework and can be accessed at https://osf.io/nw5hv (Lai & Le Forestier, 2024). Preregistrations for Study 1 and 2's designs and analysis plans are available at https://osf.io/vxtn8/ and https://osf.io/b3txc/.

Results

For each of the 15 conditions with an intervention, we conducted *t* tests comparing the intervention against the control condition. We also preregistered *t* tests comparing intervention conditions against each other. Those results are reported in the Supplemental Materials and additional online material at https://osf.io/tdr2g. We use Cohen's *d*s corresponding to those *t* tests to compare the sensitivity of measures with the various interventions.

Baseline Control Condition Descriptives

In the control condition, we observed anti-fat/pro-thin bias in all relative implicit measures. However, the magnitude of the anti-fat/pro-thin bias differed greatly. Bias was strongest on the IAT by a large margin, $M = 0.52$, $SD = 0.39$, $t(765) = 36.82$, $p < .001$, $d = 1.33$. The BIAT had the second largest effect, $M = 0.25$, $SD = 0.47$, $t(748) = 14.44$, $p < .001$, $d = 0.53$; followed by the EPT, $M = 0.12$, $SD = 0.28$, $t(795) = 12.00$, $p < .001$, $d = 0.41$; and the AMP, $M = 0.05$, $SD = 0.20$, $t(626) = 6.88$, $p < .001$, $d = 0.27$. We also observed anti-fat/pro-thin bias on the measure of explicit attitudes, $M = 0.50$, $SD = 0.90$, $t(3,655) = 33.96$, $p < .001$, $d = 0.56$.

Primary Analyses

Effectiveness by Intervention. The emotional narrative was consistently effective at reducing implicit weight bias on three of four implicit measures (see Figure 1). Relative to control, the emotional narrative reduced implicit bias on the IAT, $M = 0.42$, $SD = 0.45$, $t(1,425.10) = 4.50$, $p < .001$, $d = -0.23$; BIAT, $M = 0.18$, $SD = 0.45$, $t(1,499.47) = 2.71$, $p < .001$, $d = -0.14$; and AMP, $M = 0.01$, $SD = 0.18$, $t(1,247.37) = 4.35$, $p < .001$, $d = -0.24$. However, it failed to reduce implicit bias on the EPT, $M = 0.12$, $SD = 0.25$, $t(1,513.97) = -0.10$, $p = .92$, $d = 0.01$.

Active evaluative conditioning was consistently effective at reducing implicit weight bias on all four implicit measures. Relative to control, active evaluative conditioning reduced implicit bias on the IAT, $M = 0.24$, $SD = 0.46$, $t(1,506.18) = 12.93$, $p < .001$, $d = -0.66$; BIAT, $M = 0.06$, $SD = 0.54$, $t(1,450.57) = 7.30$, $p < .001$, $d = -0.38$; AMP, $M = 0.01$, $SD = 0.20$, $t(1,262.59) = 4.42$, $p < .001$, $d = -0.25$; and EPT, $M = 0.04$, $SD = 0.37$, $t(1,447.94) = 4.65$, $p < .001$, $d = -0.23$.

Egalitarian goals intervention was consistently ineffective at reducing implicit bias across all four implicit measures. Relative to control, egalitarian goals did not reduce implicit bias on the IAT, $M = 0.49$, $SD = 0.39$, $t(1,401.58) = 1.13$, $p = .26$, $d = -0.06$; BIAT, $M = 0.26$, $SD = 0.46$, $t(1,378.81) = -0.53$, $p = .59$, $d = 0.03$; AMP, $M = 0.05$, $SD = 0.21$, $t(1,131.21) = 0.33$, $p = .74$, $d = -0.02$; and EPT, $M = 0.12$, $SD = 0.32$, $t(1,418.54) = 0.11$, $p = .92$, $d = 0.01$.

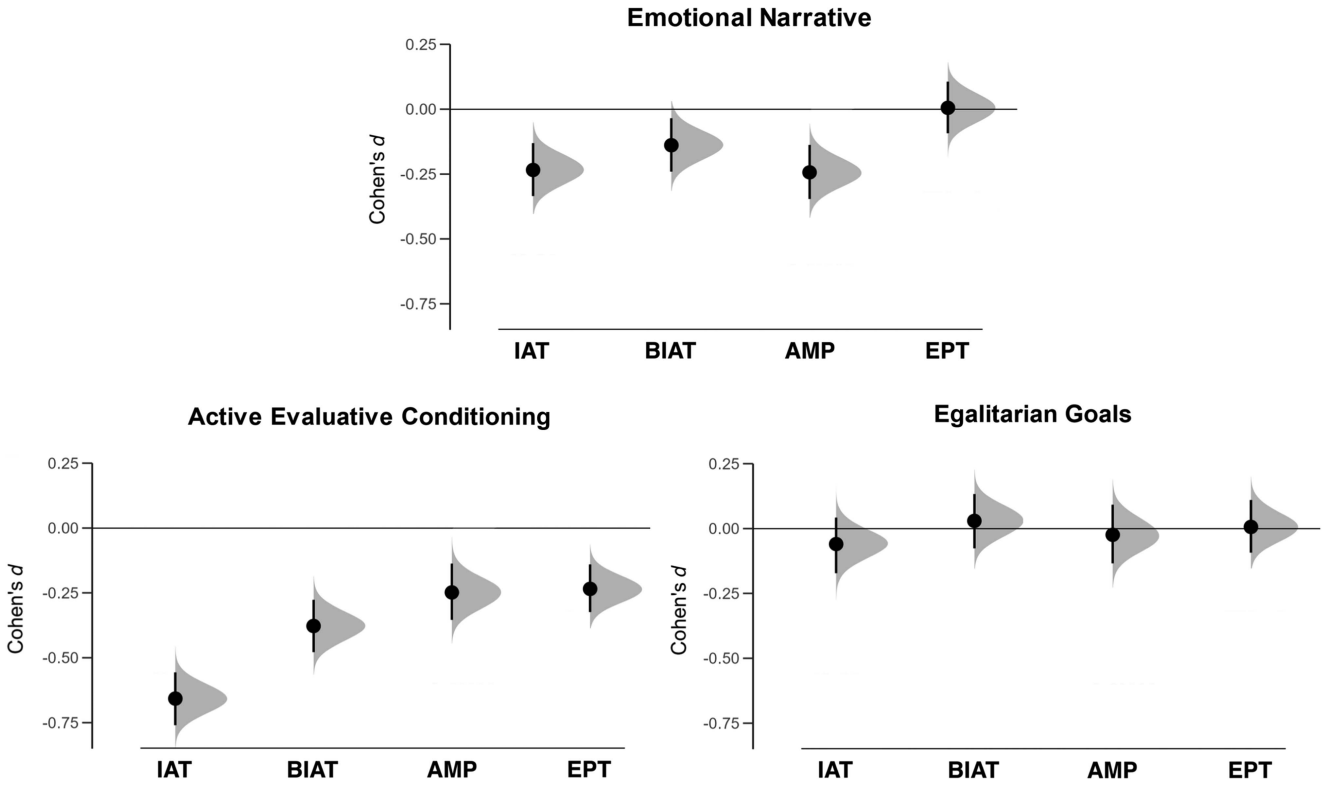
Effectiveness by Implicit Measure. Another way to characterize the results is by measure. As mentioned before, the egalitarian goals intervention failed to change implicit bias on any of the implicit measures. The IAT showed the largest and most consistent effects across emotional narrative and active evaluative conditioning, *d*s = −0.23, −0.66. The BIAT, *d*s = −0.14, −0.38, and AMP, *d*s = −0.24, −0.25, showed consistent but smaller effects across emotional narrative and active evaluative conditioning. Finally, the EPT showed the smallest and least consistent effects, with a small effect on the active evaluative conditioning, $d = -0.23$, and no observed change on the emotional narrative, $d = 0.01$.

While the IAT ($\alpha = .86$), BIAT ($\alpha = .82$), and AMP ($\alpha = .78$) showed high internal consistency, the EPT showed extremely low internal consistency ($\alpha = .24$). To examine the extent to which differences in internal consistency explained differences in intervention efficacy, we first converted Cohen's *d*s to bivariate correlations between condition (0 = control, 1 = intervention) and the implicit measure scores. Then, we corrected for attenuation by dividing the correlation by the observed reliability of each measure (Spearman, 1904). These corrected correlations allowed us to estimate the relationship between the intervention and implicit measures as if the implicit measures were measured with perfect reliability.

The corrected correlations did not substantially deviate from the bivariate correlations for the IAT, BIAT, and AMP. The corrected correlations also did not deviate from the bivariate correlations for EPT's relationship to the two conditions for which a null effect was observed (emotional narrative and egalitarian goals). However, correcting for attenuation did increase the relationship between EPT and active evaluative conditioning from $r = -.12$ to $r = -.24$. These findings suggest that differences in internal consistency explained why active evaluative conditioning had a smaller effect on the EPT but did not explain other differences (e.g., why the EPT did not detect changes in the emotional narrative).

Figure 1

Effectiveness of Interventions on Implicit Biases Relative to the Control Condition for Each of the Relative Implicit Measures in Study 1



Note. More negative Cohen's *d*s reflected greater reduction in implicit biases relative to control. Point estimates are depicted as dots. Error bars are 95% confidence intervals, and distributions are bootstrapped sampling distributions of the effects estimated using the dabestr R package. IAT = Implicit Association Test; BIAT = Brief IAT; AMP = Affect Misattribution Procedure; EPT = Evaluative Priming Task.

Secondary Analyses

Intervention-Specific Mechanisms. Answers on the mechanism questions were related to weaker implicit weight bias in the active evaluative conditioning and egalitarian goals condition, but not the emotional narrative condition. Participants in the emotional narrative condition were moderately affected by the story on average (M = 2.93/5.00, SD = 0.97). However, variations in emotions experienced were not related to implicit bias (rs = .00, −.02, .08). About half (51.1%) of participants in the active evaluative conditioning condition got all eight contingency memory questions correct, 27.8% got at least one correct, and 21.0% got none of them correct. The number of correct questions was negatively related to implicit bias on the IAT, $r(775) = -.23$, $p < .001$; BIAT, $r(738) = -.11$, $p = .0019$; EPT, $r(780) = -.08$, $p = .035$; and AMP, $r(637) = -.17$, $p < .001$. Participants in the egalitarian goals condition generally reported high goals to be egalitarian toward fat people (M = 4.21/5.00, SD = 0.83). Stronger egalitarian goals were negatively related to implicit bias on the IAT, $r(657) = -.13$, and AMP, $r(542) = -.15$, but not the BIAT, $r(644) = -.06$, $p = .12$. These findings should be taken with caution as there was no comparison to a matched control condition.

Single-Target Biases. We found that people held positive absolute implicit attitudes toward both thin and fat people, with stronger positivity toward thin people (d_{mean} = 0.53) compared with fat people (d_{mean} = 0.18) in the control condition. Pro-thin biases were moderate to large in the control condition on the EPT, M = 0.17, SD = 0.20, $t(795) = 24.09$, $p < .001$, $d = 0.85$, and the AMP, M = 0.61, SD = 0.21, $t(626) = 13.16$, $p < .001$, $d = 0.53$. By comparison, pro-fat biases were positive but weak in the control condition were positive but weak on the ST-IAT, M = 0.02, SD = 0.29, $t(815) = 2.14$, $p = .032$, $d = 0.07$; AMP, M = 0.56, SD = 0.22, $t(626) = 6.37$, $p < .001$, $d = 0.25$; and EPT, M = 0.05, SD = 0.26, $t(795) = 5.90$, $p < .001$, $d = 0.21$.

The patterns observed in implicit positivity toward thin and fat people were remarkably similar to patterns on the explicit feeling thermometer, which found strong positivity toward thin people (d = 0.67) and weak positivity toward fat people (d = 0.25). Given that 50% of the sample identified as fat and only 13% identified as thin, these findings do not align with an ingroup favoritism account. Rather, they seem to align with system justification perspectives that theorize people are motivated to justify the existing social order even when it results in outgroup favoritism (Jost et al., 2004).

Measures of single-target biases were less sensitive to change than measures of relative preferences. None of the three interventions changed ST-IAT scores relative to the control condition, ps > .05. Interventions changed AMP scores relative to the control condition one out of three times for pro-thin biases and one out of three times for pro-fat biases. Specifically, active evaluative conditioning reduced positivity about thin people on the AMP, $t(1{,}259.61) = 2.37$, $p = .018$,

$d = -0.13$, and the emotional narrative increased positivity toward fat people on the AMP, $t(1,252.29) = 2.91$, $p = .0037$, $d = 0.16$. Only the active evaluative conditioning intervention changed EPT scores, reducing positivity toward thin people, $t(1,565.19) = 2.55$, $p = .011$, $d = -0.13$, and increasing positivity toward fat people, $t(1,494.31) = 3.44$, $p < .001$, $d = 0.17$.

Explicit Weight Attitudes (Exploratory Analyses). Although we did not expect that explicit weight attitudes would be affected by our interventions that were selectively designed to change implicit attitudes, we conducted exploratory analyses examining explicit weight attitudes. Overall, implicit and explicit weight biases were weakly positively related on all measures except for the EPT: IAT, $r(2,871) = .17$, $p < .001$; BIAT, $r(2,823) = .19$, $p < .001$; EPT, $r(2,905) = .01$, $p = .73$; and AMP, $r(2,421) = .17$, $p < .001$. We found that the emotional narrative slightly reduced explicit pro-thin/anti-fat attitudes compared with control, $M = 0.39$, $SD = 0.86$, $t(7,173.85) = 5.30$, $p < .001$, $d = -0.12$. Active evaluative conditioning did not affect self-reported attitudes compared with control, $M = 0.49$, $SD = 0.88$, $t(7,260.83) = 0.82$, $p = .41$, $d = -0.02$. Ironically, egalitarian goals slightly increased explicit pro-thin/anti-fat attitudes compared with control, $M = 0.55$, $SD = 0.90$, $t(6,716.69) = -2.33$, $p = .02$, $d = 0.06$.

Discussion

In Study 1, we found that two of three interventions reduced implicit weight biases across the implicit measures. The emotional narrative and active evaluative conditioning interventions were both effective at reducing implicit weight biases across most of the implicit measures, with effects being particularly strong for active evaluative conditioning. By contrast, the egalitarian goals intervention was ineffective at changing implicit weight biases.

Examining intervention efficacy by implicit measure, one measure stood out: the EPT. Intervention effects on the EPT were smaller than all other measures. The EPT's lower measurement reliability was a partial explanation for why intervention effects were weaker but did not fully account for why the EPT was less sensitive to intervention than other measures. We also examined the effect of interventions on single-target biases and found much less consistent evidence for change on those measures. Finally, we found that interventions were generally ineffective at changing explicit attitudes.

We had the opportunity to examine the theoretical mechanisms that could explain intervention efficacy. We found that greater contingency memory in the active evaluative conditioning condition and stronger self-reported egalitarian goals were related to reduced implicit biases in the egalitarian goals condition. By contrast, we did not find that the intensity of emotional experience was related to reduced implicit biases in the emotional narrative condition. However, we hesitate to draw strong conclusions from those findings due to the lack of matched control conditions. The observed correlations could be attributed to individual differences (e.g., more prejudiced people paid less attention to the intervention). We explored this further in Study 2.

Study 2

Study 2 expanded on Study 1 by testing two new interventions. Although the active evaluative conditioning intervention was effective, the active categorization approach is unconventional for evaluative conditioning research (Gast et al., 2020). By contrast, passive exposure to stimuli is a more common approach to evaluative conditioning that generates reliably small changes in implicit biases (Kurdi et al., 2023). Passive evaluative conditioning approaches also give greater experimental isolation of mental tracking of co-occurrence. So, we replaced the active evaluative conditioning intervention with a passive evaluative conditioning intervention in Study 2. We also tested a newer approach to intervention that recent accounts of implicit social cognition indicate is more influential for implicit attitude change: evaluative statements. Directly telling participants how to think differently about the relationships between groups and valence has been shown to be even more effective than passive evaluative conditioning procedures (Kurdi & Banaji, 2017, 2019; Kurdi et al., 2023). This intervention replaced the ineffective egalitarian goals intervention.

Study 2 also introduced matched control conditions for each intervention condition to better measure the theoretical mechanisms underlying interventions. These matched control conditions were identical to the intervention conditions but were about age instead of weight. If a measured mechanism was relevant for implicit attitude change (e.g., contingency memory), then we would find that the measured mechanism would be more strongly related to implicit weight biases in the intervention conditions that targeted weight prejudice than the corresponding matched control condition that targeted age prejudice. Through these analyses, we would be able to examine the mechanisms underlying intervention efficacy and test whether those mechanisms matter more for some implicit measures than others.

Method

Participants

Our final analytic sample included 12,734 participants from the research platform Project Implicit (https://implicit.harvard.edu). The final sample was 68% female and 32% male. The mean age was 41.9, and 71% of participants identified as White, 7% as Black, 6% as Asian, 9% as Hispanic/Latino, 0.3% as Native Hawaiian or Pacific Islander, 0.3% as American Indian/Alaska Native, 4% as multiracial, and 2% as other or unknown. For the highest level of education, 2% had less than a high school degree, 14% had a high school degree, 42% had an undergraduate degree, and 42% had a postgraduate degree. Politically, 71% identified as liberal, 2,824 identified as moderate, and 15% identified as conservative.

We originally planned to collect an average of 291.33 participants per matched control condition and an average of 582 participants per intervention condition, totaling 13,637 participants. We selected this unequal allocation ratio as most analyses would be conducted with the three matched control conditions collapsed together into a combined control condition that would be 50% larger than each intervention condition. Our power analyses indicated this sample size would grant us 80% power to detect mean differences of $d = 0.15$ between an intervention condition and the combined control condition (assuming 4.1% data excluded, consistent with Study 1). Eighteen thousand seven hundred eighty-five American residents or citizens consented to take the study, of which 13,248 (70.5%) completed the study. We then excluded 514 participants (3.9%) who violated implicit measure exclusion criteria for our final sample of 12,734. This consisted of 8/2613 (0.3%) participants who took the

IAT, 12/2557 (0.4%) participants who took the BIAT, 20/2663 (0.8%) participants who took the EPT, 485/2959 (16.4%) participants who took the AMP, and 5/2706 (0.1%) participants who took the ST-IAT. We followed the exclusion criteria used by Bar-Anan and Nosek's (2014) study of the psychometrics of seven implicit measures.

Procedure

The structure of Study 2 was similar to Study 1. Participants were randomly assigned to one of 30 conditions in a 6 (intervention: emotional narrative, passive evaluative conditioning, evaluative statement, matched control: emotional narrative, matched control: passive evaluative conditioning, matched control: evaluative statement) × 5 (implicit measure: IAT, ST-IAT, BIAT, AMP, EPT) design. Random assignment for intervention condition was skewed such that participants had a 22% chance to receive any of the three intervention conditions (66% total) and an 11% chance to receive any of the three matched control conditions (33% total). After taking an intervention or matched control, participants completed the implicit measure they were randomly assigned to. Finally, they completed self-report measures measure of explicit weight attitudes, potential intervention-specific mechanisms, and several additional exploratory measures (i.e., perceived control over body weight and indices of subjective and objective weight status).

Beyond the change in experimental conditions, we also (a) replaced the category labels "good" and "bad" on the AMP with "pleasant" and "unpleasant" to match standard conventions at the expense of more mismatch with the other implicit measures and (b) added new moderator questions for the passive evaluative conditioning, evaluative statement, and matched control conditions.

Passive Evaluative Conditioning (Adapted From Kurdi & Banaji, 2017). Participants were exposed to stimulus pairings reinforcing that fat = good and thin = bad. They first received general instructions about the learning task and were then presented with 37 conditioning trials. In each trial, a fat person was paired with a positive image, or a thin person was paired with a negative image. Participants passively observed the pairings and did not take any action on each trial. Each pairing was presented for 2,500 ms with an intertrial interval of 1,000 ms. Unlike in Study 1's active evaluative conditioning intervention, the stimuli were not the same as the ones used in the implicit measures. The fat and thin faces were age-diverse stimuli taken from several face databases, and the valenced images came from Kurdi and Banaji's (2017) studies.

Evaluative Statement (Adapted From Kurdi & Banaji, 2017). Participants were given a verbal instruction reinforcing that fat = good and thin = bad. They received the same general instructions about the learning task as the passive evaluative conditioning intervention but were informed that they would learn that fat people would always be paired with good things and thin people would always be paired with bad things. Participants in this intervention did not receive any conditioning trials before taking the dependent measures. They were offered an opportunity to go through the conditioning trials at the end of the study after all other measures had been completed.

Matched Control Conditions. All three of the matched control conditions were identical to the intervention conditions, except that all verbal references to "fat" and "thin" people were replaced with references to "old" and "young" people. The face images in the passive evaluative conditioning and evaluative statement matched control conditions were identical to the intervention conditions, but the stimulus pairings focused on "old + good things" versus "young + bad things."

Intervention-Specific Mechanisms. The questions for the emotional narrative condition were the same as in Study 1.

Participants in the passive evaluative conditioning and matched control condition were asked questions to test their memory for the contingencies between stimuli. They were shown each of the 10 images in the task (thin/fat people in passive evaluative conditioning; old/young people in the matched control) and asked to select what type of words they were paired with: "positive words," "negative words," "neutral words," or "unsure." For analyses, we compared participants who got all questions correct (N = 1,407) against participants who got at least one question wrong (N = 2,881).

Participants in the evaluative statement condition were asked, "What were you told you would see at the beginning of the study? Check all that apply." and given four response options: "Fat People will always be paired with Good Things," "Fat People will always be paired with Bad Things," "Thin People will always be paired with Good Things," and "Thin People will always be paired with Bad Things." Participants in the corresponding matched control condition saw the same question, but with references to "fat" and "thin" people replaced with references to "old" and "young people." For analyses, we compared participants who got the question correct (N = 3,127) against participants who got the question wrong (N = 927).

Results

We combined the three matched control conditions for our primary analyses. For simplicity, we will refer to these combined conditions as the "control condition." The preregistered analyses were nearly identical to Study 1 otherwise, with additional preregistered analyses comparing intervention conditions against each other and analyses comparing intervention conditions against individual matched control conditions reported in the Supplemental Materials and additional online material at https://osf.io/tdr2g.

Matched Control Condition Descriptives

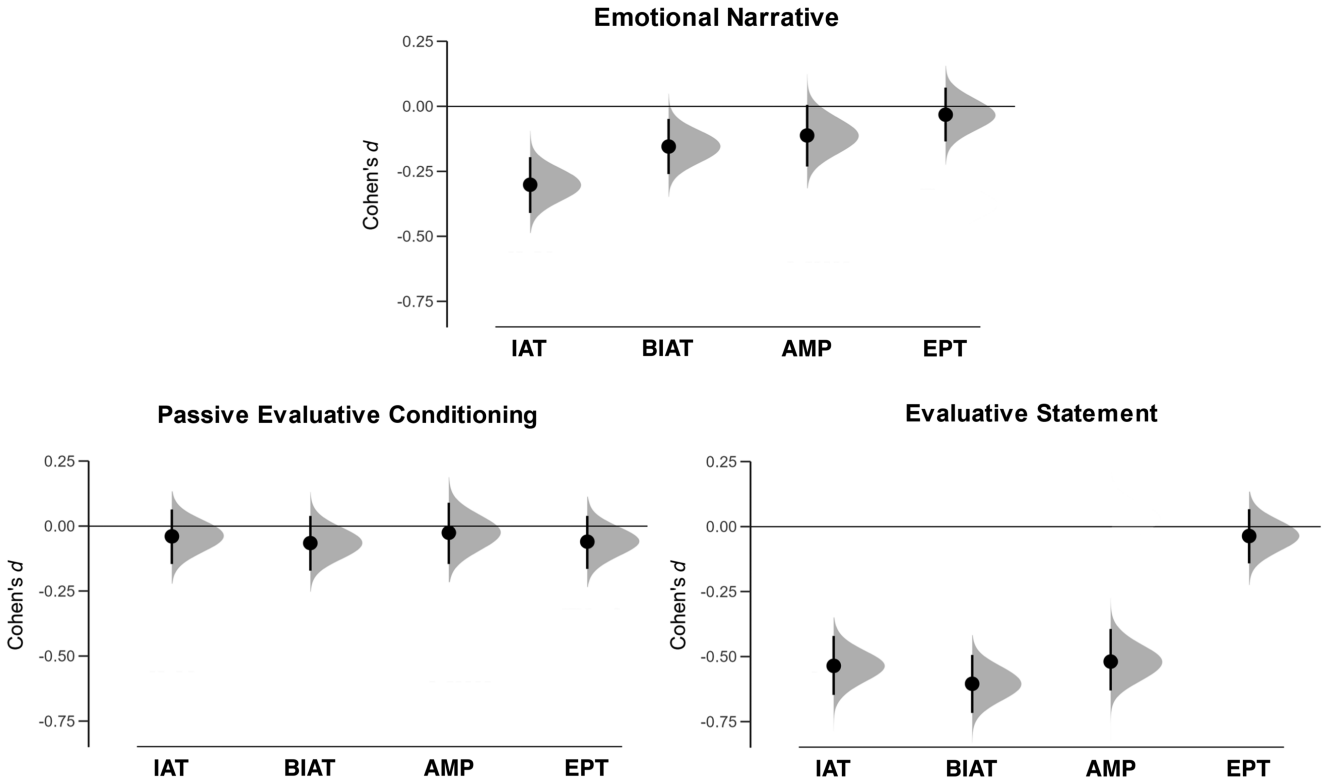
In the control condition, we observed anti-fat/pro-thin bias in all relative implicit measures. However, the magnitude of the anti-fat/pro-thin bias differed greatly. Bias was strongest on the IAT by a large margin, $M = 0.47$, $SD = 0.40$, $t(838) = 34.35$, $p < .001$, $d = 1.19$. The BIAT had the second largest effect, $M = 0.23$, $SD = 0.47$, $t(827) = 14.27$, $p < .001$, $d = 0.50$; followed by the EPT, $M = 0.07$, $SD = 0.26$, $t(918) = 7.84$, $p < .001$, $d = 0.26$; and the AMP, $M = 0.05$, $SD = 0.19$, $t(660) = 6.98$, $p < .001$, $d = 0.27$. We also observed anti-fat/pro-thin bias on the measure of explicit attitudes, $M = 0.49$, $SD = 0.87$, $t(2{,}926) = 30.26$, $p < .001$, $d = 0.56$.

Primary Analyses

Effectiveness by Intervention. The emotional narrative reduced implicit weight bias on two of four implicit measures (see Figure 2). Relative to control, the emotional narrative reduced implicit bias on the IAT, $M = 0.35$, $SD = 0.43$, $t(1{,}144.28) = 5.46$, $p < .001$, $d = -0.30$, and BIAT, $M = 0.16$, $SD = 0.48$, $t(1{,}246.13) = 2.85$, $p < .001$, $d = -0.15$. However, it failed to reduce implicit bias on the

Figure 2

Effectiveness of Interventions on Implicit Biases Relative to the Control Condition for Each of the Relative Implicit Measures in Study 2



Note. More negative Cohen's *d*s reflected greater reduction in implicit biases relative to control. Point estimates are depicted as dots. Error bars are 95% confidence intervals, and distributions are bootstrapped sampling distributions of the effects estimated using the dabestr R package. IAT = Implicit Association Test; BIAT = Brief IAT; AMP = Affect Misattribution Procedure; EPT = Evaluative Priming Task.

AMP, $M = 0.03$, $SD = 0.18$, $t(1{,}017.31) = 1.86$, $p = .063$, $d = -0.11$, and the EPT, $M = 0.06$, $SD = 0.26$, $t(1{,}220.36) = 0.60$, $p = .55$, $d = -0.03$.

Passive evaluative conditioning failed to reduce implicit weight bias on any of the implicit measures. Relative to control, passive evaluative conditioning failed to reduce implicit bias on the IAT, $M = 0.46$, $SD = 0.44$, $t(1{,}237.31) = 0.73$, $p = .47$, $d = -0.04$; BIAT, $M = 0.20$, $SD = 0.48$, $t(1{,}243.8) = 1.21$, $p = .23$, $d = -0.07$; AMP, $M = 0.05$, $SD = 0.19$, $t(1{,}002.64) = 0.42$, $p = .67$, $d = -0.03$; and EPT, $M = 0.05$, $SD = 0.25$, $t(1{,}284.67) = 1.15$, $p = .25$, $d = -0.06$.

The evaluative statement reduced implicit weight bias on three of four implicit measures. Relative to control, the evaluative statement reduced implicit bias on the IAT, $M = 0.24$, $SD = 0.48$, $t(1{,}122.71) = 9.67$, $p < .001$, $d = -0.54$; BIAT, $M = -0.06$, $SD = 0.50$, $t(1{,}138.17) = 10.89$, $p < .001$, $d = -0.60$; and AMP, $M = -0.08$, $SD = 0.32$, $t(699.53) = 7.89$, $p < .001$, $d = -0.52$. However, it failed to reduce implicit bias on the EPT, $M = 0.06$, $SD = 0.25$, $t(1{,}121.68) = 0.67$, $p = .51$, $d = -0.04$.

Effectiveness by Implicit Measure. Another way to characterize the results is by measure. As recently discussed, none of the implicit measures were affected by passive evaluative conditioning. The IAT showed the largest and most consistent effects across the emotional narrative and evaluative statement conditions, *d*s = −0.30, −0.54. The BIAT also showed consistent effects across those two conditions, *d*s = −0.15, −0.60. Only the evaluative statement reduced AMP scores, $d = -0.52$. Finally, the EPT was not affect by any intervention, *d*s = −0.06, −0.03, −0.04.

While the IAT (α = .86), BIAT (α = .82), and AMP (α = .84) showed high internal consistency, the EPT showed remarkably low internal consistency (α = −.11).[3] As in Study 1, we computed correlations that were corrected for attenuation to estimate the relationship between the intervention and implicit measures as if the implicit measures were measured with perfect reliability.

The corrected correlations did not substantially deviate from the bivariate correlations for the IAT, BIAT, and AMP. As the EPT's internal consistency was unexpectedly negative, it was not possible to compute corrected correlations. To compute a rough but optimistic estimate, we calculated corrected correlations that substituted EPT's observed internal consistency (α = −.11) with EPT's internal consistency from a recent meta-analysis (α = .53, Greenwald & Lai, 2020). Correcting for attenuation using the

[3] As the internal consistency was extremely low and inconsistent with Study 1's reliability, we were concerned that this result was due to a coding, data processing, or analysis error. We traced the results back through our data processing pipeline to the raw data and compared our code with the code from Study 1. We could not find any discrepancies between the two studies, suggesting that the extremely low internal consistency could not be attributed to an error unique to Study 2.

meta-analytic internal consistency did not substantially improve the relationship between EPT and the intervention conditions. That suggests that the lack of internal consistency did not completely explain for EPT's lack of sensitivity to change.

Secondary Analyses

Intervention-Specific Mechanisms. If experienced emotions in the emotional narrative condition and contingency memory in the passive evaluative conditioning and evaluative statement conditions were important for intervention effects, then implicit weight biases should be more strongly related to experienced emotions and contingency memory in the intervention conditions that were about fat/thin people than in the matched control conditions where the procedures were about old/young people. To investigate this, we examined whether experienced emotions and contingency memory moderated intervention effects across each of the five implicit measures when compared with their matched control conditions. They did not in all 15 of the moderation analyses we conducted.

Single-Target Biases. In line with relative measures of implicit preference, pro-thin biases were moderate to large in the control condition on the EPT, $M = 0.23$, $SD = 0.22$, $t(918) = 30.96$, $p < .001$, $d = 1.02$, and the AMP, $M = 0.64$, $SD = 0.20$, $t(660) = 17.63$, $p < .001$, $d = 0.66$. By comparison, pro-fat biases in the control condition were positive but weaker than pro-thin biases on the ST-IAT, $M = 0.05$, $SD = 0.32$, $t(883) = 4.38$, $p < .001$, $d = 0.15$; EPT, $M = 0.16$, $SD = 0.22$, $t(918) = 22.02$, $p < .001$, $d = 0.73$; and AMP, $M = 0.59$, $SD = 0.22$, $t(660) = 9.95$, $p < .001$, $d = 0.39$. As in Study 1, these findings align with system justification perspectives theorizing that people hold status-quo reinforcing attitudes toward high-status groups (Jost et al., 2004).

Only the evaluative statement consistently changed measures of single-target biases. The evaluative statement reduced pro-thin biases on the EPT, $t(1{,}111.16) = 3.90$, $p < .001$, $d = -0.21$, and the AMP, $t(866.74) = 6.15$, $p < .001$, $d = -0.39$. The evaluative statement also increased pro-fat biases on the EPT, $t(1{,}104.58) = 3.13$, $p < .001$, $d = 0.22$, and the AMP $t(969.25) = 3.40$, $p < .001$, $d = 0.21$. However, it failed to change ST-IAT scores, $t(1{,}373.18) = -0.47$, $p < .001$, $d = -0.02$. The emotional narrative failed to change any of the five measures of single-target biases, which meant that it failed to directly replicate changes in pro-thin and pro-fat biases observed in Study 1. Passive evaluative conditioning slightly reduced pro-thin biases on the EPT, $M = 0.20$, $SD = 0.23$, $t(1{,}207.28) = 2.46$, $p = .014$, $d = -0.13$, but did not affect any other measure.

Explicit Weight Attitudes (Exploratory Analyses). As in Study 1, we conducted exploratory analyses examining the impact of interventions on explicit weight attitudes. Implicit and explicit weight biases were weakly positively related on all measures: IAT, $r(1{,}869) = .21$, $p < .001$; BIAT, $r(1{,}779) = .15$, $p < .001$; EPT, $r(1{,}857) = .05$, $p = .026$; and AMP, $r(1{,}445) = .17$, $p < .001$. The emotional narrative slightly reduced explicit pro-thin/anti-fat attitudes compared with control, $M = 0.41$, $SD = 0.87$, $t(6{,}007.68) = 3.47$, $p < .001$, $d = -0.09$. Passive evaluative conditioning did not affect self-reported attitudes compared with control, $M = 0.51$, $SD = 0.88$, $t(5{,}927.14) = -0.95$, $p = .34$, $d = 0.02$. Evaluative statement also failed to change explicit pro-thin/anti-fat attitudes compared with control, $M = 0.46$, $SD = 0.95$, $t(5{,}455.17) = 1.08$, $p = .28$, $d = 0.03$.

General Discussion

In two large experiments totaling 27,114 people, we tested the efficacy of five interventions across five implicit measures. Our megastudies provide the first large-scale comparative evidence of interventions for reducing implicit weight prejudice, finding that implicit weight prejudice was consistently strong but amenable to change. The most effective intervention for reducing pro-thin/anti-fat biases across the implicit measures was the evaluative statement ($d_{average} = -0.43$), followed by active evaluative conditioning ($d_{average} = -0.38$) and emotional narrative ($d_{average} = -0.15$). By contrast, passive evaluative conditioning ($d_{average} = -0.05$) and egalitarian goals ($d_{average} = -0.03$) were consistently ineffective for reducing implicit biases. When comparing implicit measures, we found that the IAT, BIAT, and AMP were consistently sensitive to change. Effective approaches to change one of these implicit measures often generalized to the other two measures. By contrast, the EPT, the ST-IAT, and other approaches to measuring single-target biases were often insensitive to change.

Implications for Prejudice Reduction

The relative effectiveness of interventions is consistent with a recent meta-analysis establishing that propositional information about how concepts are related to each other is dominant for implicit attitude change generally, far outweighing associative information about how stimuli co-occur together (Kurdi et al., 2023). Evaluative statement and active evaluative conditioning yielded the strongest effects, perhaps because these interventions required people to consciously practice a rule linking fat people with good things and thin people with bad things. By contrast, the emotional narrative highlighted the relationships between concepts more indirectly by hinging on the assumption that people will generalize an emotional experience with a heroic fat man and villainous thin man to attitudes about fat and thin people in general.

Finally, the two ineffective interventions were the least direct in communicating the causal relations between key concepts. Passive evaluative conditioning exclusively relied on associative information that paired fat people with good things and thin people with bad things without any explanation about how those concepts are related to each other, which is known to generate smaller effects than information about how concepts are related to each other (Kurdi et al., 2023). We suspect that distinction is the most relevant one for why passive evaluative conditioning yielded smaller effects than active evaluative conditioning. Passive evaluative conditioning may have also been less effective as it did not use stimuli matched to the implicit measures' stimuli, whereas active evaluative conditioning did. Rather than changing attitudes to the social category of fat and thin people, then, active evaluative conditioning may have changed attitudes to the individual fat and thin people represented in the stimuli. We cannot rule this possibility out with our experimental design but give more weight to the propositional versus evaluative information distinction given the larger body of evidence underlying that account of implicit attitude change. We are also aware of two studies that experimentally tested whether stimuli matched to a race IAT moderated the effects of two evaluative conditioning interventions (Lai et al., 2016). Both studies found that it did not, which speaks against the possibility of matched stimuli mattering.

Egalitarian goals, the other ineffective intervention, relied on the assumption that feeling that one has failed to be egalitarian to fat people would translate into inhibition of anti-fat prejudice. These findings did not conceptually replicate the results of the original intervention studies of Moskowitz and Li (2011), which found that the intervention inhibited racial stereotyping on implicit measures modeled after the lexical decision task (Wittenbrink et al., 1997) and the Stroop task (Stroop, 1935). One possibility is that people were less motivated to inhibit their anti-fat prejudices than their racial prejudices. While that is plausible, a study by Lai et al. (2014) had also found that the intervention failed to shift racial attitudes measured by the IAT. That suggests the failures to replicate extend beyond weight attitudes to racial attitudes as well. Another possibility is that the original results were a false positive, suggesting that this manipulation of egalitarian goals is generally ineffective for automatically inhibiting anti-fat prejudice. While the best way to test this possibility would be a direct replication, the relative sample sizes give a reason to assign greater confidence to the more recent studies' larger samples (Study 1 egalitarian goals vs. control $N = 5{,}492$ and Lai et al., 2014, $N = 339$ vs. Moskowitz & Li, 2011, Ns = 42, 38, 38, 36). Perhaps another manipulation of egalitarian goals may have yielded stronger effects (e.g., N. H. Mann & Kawakami, 2012). Otherwise, egalitarian goals may just be ineffective at inhibiting prejudice automatically. Egalitarianism could also be difficult to temporarily induce at a strength necessary to shift implicit attitudes and may be better understood as a chronic disposition (Moskowitz & Ignarri, 2009; Perry et al., 2015) or a tendency that unfolds over longer periods of time (Monteith, 1993).

Our studies suggest several practical ideas for implicit prejudice interventions in real-world contexts. First, mere exposure to people who defy one's prejudices (i.e., liked fat people or disliked thin people) is not sufficient to change implicit weight prejudice. As with research on counterattitudinal exemplars in implicit racial attitudes (Joy-Gaba & Nosek, 2010; Kurdi et al., 2024), the power of a counterattitudinal individual to shift implicit prejudices arises when people consciously appraise that individual in terms of valence and the social category of fat or thin people. That means changes in media representation may not be sufficient. Rather, changes in media representation would ideally be paired with messaging that causes people to think of the social category of fat people in a positive light and/or the category of thin people in a negative light.

Second, our successful interventions went beyond targeting anti-fat animus to target pro-thin favoritism as well. These interventions were inspired by studies finding that implicit racial prejudice interventions are more influential in reducing implicit pro-White bias than implicit anti-Black bias (Calanchini et al., 2021) and the insight that prejudice and discrimination often arise from a *relative* difference in evaluation (Brewer, 1999; Greenwald & Pettigrew, 2014). All that said, reducing positivity toward a group is morally fraught. Real-world interventions may consider targeting pro-thin favoritism as well but will have to contend with political frames that prevent backlash (e.g., "thin privilege") and the possibility that doing so may not lead to better treatment for fat people in an absolute sense.

Study 2's sophisticated matched control design afforded the ability to investigate potential mechanisms directly. We found that experienced emotions in the emotional narrative condition was not differentially related to implicit biases compared with its matched control condition. We suspect that is because the emotional narrative's effectiveness can primarily be attributed to rule-based learning rather than the intensity of emotional experience. We also found that contingency memory in the evaluative statement and passive evaluative conditioning conditions was not differentially related to implicit biases compared with that of matched control conditions. With hindsight, we suspect that the lack of moderation could be due to inherent limitations of contingency memory measures (Gawronski & Walther, 2012). The lack of moderation could be attributed to unconscious attitude change during encoding or just a failure to retrieve the relevant memories in self-report. Future research could use more refined experimental designs to isolate the mechanisms that explain how the interventions changed implicit biases.

We also found that interventions that were effective in reducing implicit weight biases did so without reducing explicit weight prejudice. This independence in changing implicit and explicit biases held regardless of what implicit measure we used. One potential reason for the lack of corresponding explicit prejudice change is the research design. The interventions we tested were selected based on past studies finding these approaches were particularly effective for changing implicit biases. If we selected interventions that were more known for changing explicit biases (e.g., controllability of weight) or prioritized multifaceted interventions that could target implicit and explicit biases simultaneously, we might have found stronger evidence for explicit prejudice change.

Implications for Implicit Measures

Our evidence suggests that the common practice of generalizing claims about bias change on one implicit measure to how biases change on other implicit measures is often (but not always) justifiable. We found that conclusions about weight bias reduction generally held across implicit measures assessing relative preferences between thin and fat people. We found that intervention effects were consistent on implicit measures in 20 out of 24 comparisons compared with the control condition.

This consistency is notable given that implicit measures differed widely in their procedures for invoking implicit biases (e.g., priming, speeded categorization), the outcomes they used (e.g., reaction time, evaluative judgments), and the forms of automaticity they were designed to capture (e.g., efficiency, controllability, unconscious processing). We had also theorized that implicit measures could differ in their sensitivity to different sources of change (e.g., affective vs. cognitive appeals; associative vs. propositional information, responsiveness to goals) but did not find evidence for it. Notably, our direct experimental findings contrast with recent meta-analytic evidence finding that studies with the AMP had larger propositional learning effects than the studies using the EPT- or IAT-based measures (Kurdi et al., 2023).

However, one relative measure of implicit biases stood out as worse than the others: the EPT. Of the four comparisons that were inconsistent with other measures, three comparisons involved the EPT finding null effects of an intervention that was otherwise effective on the other measures. While follow-up analyses suggest that the EPT's low reliability may have played a small part in explaining differences in intervention effects, our analyses suggested that the null effects were not primarily attributable to measurement reliability. The EPT's lack of reactivity to interventions converges with Bar-Anan and Nosek's (2014) study finding that the EPT had

the worst measurement reliability, weakest correlation with other implicit measures, and weakest correlation with explicit measures out of seven implicit measures studied. Although the EPT is uniquely tuned to assess spontaneous evaluations of individual stimuli, we suspect that the EPT's poor measurement properties and insensitivity to psychological change bar it from being a dependable tool outside of studies with very large samples.

Finally, measures assessing single-target biases about thin and fat people were less sensitive to change. We speculate that this is because all of the interventions that were effective at reducing implicit biases did so by targeting both pro-thin and anti-fat attitudes simultaneously. Targeting both attitudes was consistent with prior studies finding that the interventions that are most effective for reducing implicit racial bias on the IAT often targeted both favoritism toward the advantaged groups and disliking of stigmatized groups. Interventions that only targeted disliking of the stigmatized group (i.e., Black people) were generally ineffective at changing implicit racial biases on the IAT (Calanchini et al., 2021; Lai et al., 2014). However, this two-sided approach may have increased effectiveness for relative measures at the cost of being less effective with single-target measures. Single-target measures would only be sensitive to changes in attitudes toward one group at a time (i.e., thin people or fat people). If an intervention operates by changing attitudes toward both groups simultaneously, then a single-target measure would be sensitive to changes in only one of those two attitudes. Future research could investigate whether interventions that target attitudes about thin people or fat people exclusively will be more effective for changing single-target measures.

Constraints on Generality

In the current studies, we focused on the attitudes of U.S. residents. Although anti-fat prejudices are a global phenomenon, there are meaningful cultural variations in the expression of anti-fat prejudice that could be related to intervention efficacy (Brewis, 2011). For example, the history of anti-fat stigma is more recent in many parts of the global south (Brewis et al., 2018), and countries that have higher gross domestic product (GDP) also tend to show stronger implicit pro-thin/anti-fat biases on average (Marini et al., 2013). Interventions to combat anti-fat prejudice may be interpreted differently when anti-fat prejudices reside within cultural contexts where weight stigma is less (or more) deeply ingrained (SturtzSreetharan et al., 2021), where cultural norms simultaneously reinforce fat positivity (e.g., Hardin, 2015), or where undernutrition is a public concern and thinness is more strongly stigmatized (Brewis et al., 2018).

Our studies afforded strong causal inferences about what interventions were effective at reducing implicit biases within a single session, but future research must examine the impact of interventions in the long term and their downstream consequences on behavior. These limitations are unfortunately common; most research on prejudice reduction do not assess outcomes beyond one session or behavioral outcomes (Forscher et al., 2019; Kurdi & Charlesworth, 2023; Lai et al., 2013, 2016; Paluck et al., 2021). Our studies on which interventions work at changing prejudices in the short term are a first step toward establishing which approaches could be most effective for long-term change and behavioral change but are no substitute for direct studies following up on those approaches. Intervention approaches that are effective for long-term and behavioral change could employ additional intervention features that sustain psychological change (e.g., habit formation; Frey & Rogers, 2014) or target behavior directly (e.g., matching intervention content to behavior; Lai et al., 2023; Petty & Cacioppo, 1986).

Conclusion

The severity and pervasiveness of anti-fat prejudice and discrimination have led to calls for interventions to address them. However, the pursuit of interventions to combat anti-fat prejudice has been stymied by ineffective approaches to reducing prejudice, small sample size studies, and the lack of standardization in measurement. To that end, we adapted five interventions that had previously been effective in other intergroup domains to combat implicit weight prejudice. We found that the most effective interventions instructed people to practice an explicit rule linking fat people with good things and thin people with bad things. Interventions that were more indirect or relied on associative information were more ineffective.

These findings provide a roadmap for future investigations into combating anti-fat prejudice. Future research can scale these interventions up and investigate the long-term durability of their effects on implicit weight prejudices (Lai et al., 2016), examine their influence on downstream behaviors (Forscher et al., 2019), and incorporate them into practical applications (e.g., mass media, diversity training; Devine & Ash, 2022; Paluck et al., 2021). Our research laid the foundation for understanding what approaches for changing implicit weight prejudice will be most effective; the next challenge will be building on this understanding to assess how these approaches can inform real-world prejudice reduction.

References

Alberga, A. S., Pickering, B. J., Alix Hayden, K., Ball, G. D. C., Edwards, A., Jelinski, S., Nutter, S., Oddie, S., Sharma, A. M., & Russell-Mayhew, S. (2016). Weight bias reduction in health professionals: A systematic review. *Clinical Obesity*, *6*(3), 175–188. https://doi.org/10.1111/cob.12147

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Experimental Psychology*, *48*(2), 145–160. https://doi.org/10.1026//0949-3946.48.2.145

Bar-Anan, Y., & Nosek, B. A. (2012). Reporting intentional rating of the primes predicts priming effects in the affective misattribution procedure. *Personality and Social Psychology Bulletin*, *38*(9), 1194–1208. https://doi.org/10.1177/0146167212446835

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods*, *46*(3), 668–688. https://doi.org/10.3758/s13428-013-0410-6

Bar-Anan, Y., & Nosek, B. A. (2016). *A comparison of the sensitivity of four indirect evaluation measures to evaluative information*. https://doi.org/10.31219/osf.io/p2akj

Bar-Anan, Y., & Vianello, M. (2018). A multi-method multi-trait test of the dual-attitude perspective. *Journal of Experimental Psychology: General*, *147*(8), 1264–1272. https://doi.org/10.1037/xge0000383

Blaison, C., Imhoff, R., Hühnel, I., Hess, U., & Banse, R. (2012). The affect misattribution procedure: Hot or not? *Emotion*, *12*(2), 403–412. https://doi.org/10.1037/a0026907

Bluemke, M., & Friese, M. (2008). Reliability and validity of the Single-Target IAT (ST-IAT): Assessing automatic affect towards multiple attitude objects. *European Journal of Social Psychology*, *38*(6), 977–997. https://doi.org/10.1002/ejsp.487

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues*, *55*(3), 429–444. https://doi.org/10.1111/0022-4537.00126

Brewis, A. A. (2011). *Obesity: Cultural and biocultural perspectives*. Rutgers University Press. https://www.jstor.org/stable/j.ctt5hj0mn

Brewis, A. A., SturtzSreetharan, C., & Wutich, A. (2018). Obesity stigma as a globalizing health challenge. *Globalization and Health*, *14*(1), Article 20. https://doi.org/10.1186/s12992-018-0337-x

Brochu, P. M., & Esses, V. M. (2011). What's in a name? The effects of the labels "fat" versus "overweight" on weight bias. *Journal of Applied Social Psychology*, *41*(8), 1981–2008. https://doi.org/10.1111/j.1559-1816.2011.00786.x

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), 796–818. https://doi.org/10.1037/pspi0000339

Campos-Vazquez, R. M., & Gonzalez, E. (2020). Obesity and hiring discrimination. *Economics and Human Biology*, *37*, Article 100850. https://doi.org/10.1016/j.ehb.2020.100850

Charlesworth, T. E. S., & Banaji, M. R. (2019). Patterns of implicit and explicit attitudes: I. Long-term change and stability from 2007 to 2016. *Psychological Science*, *30*(2), 174–192. https://doi.org/10.1177/0956797618813087

Cone, J., Mann, T. C., & Ferguson, M. J. (2017). Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. In J. M. Olson (Ed.), *Advances in experimental social psychology* (Vol. 56, pp. 131–199). Academic Press. https://doi.org/10.1016/bs.aesp.2017.03.001

Crandall, C. S. (1994). Prejudice against fat people: Ideology and self-interest. *Journal of Personality and Social Psychology*, *66*(5), 882–894. https://doi.org/10.1037/0022-3514.66.5.882

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, *82*(3), 359–378. https://doi.org/10.1037/0022-3514.82.3.359

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, *12*(2), 163–170. https://doi.org/10.1111/1467-9280.00328

Czellar, S. (2006). Self-presentational effects in the Implicit Association Test. *Journal of Consumer Psychology*, *16*(1), 92–100. https://doi.org/10.1207/s15327663jcp1601_11

Daníelsdóttir, S., O'Brien, K. S., & Ciao, A. (2010). Anti-fat prejudice reduction: A review of published studies. *Obesity Facts*, *3*(1), 47–58. https://doi.org/10.1159/000277067

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

Degner, J. (2009). On the (un-)controllability of affective priming: Strategic manipulation is feasible but can possibly be prevented. *Cognition and Emotion*, *23*(2), 327–354. https://doi.org/10.1080/02699930801993924

Deutsch, R., & Gawronski, B. (2009). When the method makes a difference: Antagonistic effects on "automatic evaluations" as a function of task characteristics of the measure. *Journal of Experimental Social Psychology*, *45*(1), 101–114. https://doi.org/10.1016/j.jesp.2008.09.001

Devine, P. G., & Ash, T. L. (2022). Diversity training goals, limitations, and promise: A review of the multidisciplinary literature. *Annual Review of Psychology*, *73*(1), 403–429. https://doi.org/10.1146/annurev-psych-060221-122215

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*(6), 1013–1027. https://doi.org/10.1037/0022-3514.69.6.1013

Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the Implicit Association Tests. *Basic and Applied Social Psychology*, *27*(4), 307–316. https://doi.org/10.1207/s15324834basp2704_3

Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2019). A meta-analysis of procedures to change implicit measures. *Journal of Personality and Social Psychology*, *117*(3), 522–559. https://doi.org/10.1037/pspa0000160

Frey, E., & Rogers, T. (2014). Persistence: How treatment effects persist after interventions stop. *Policy Insights From the Behavioral and Brain Sciences*, *1*(1), 172–179. https://doi.org/10.1177/2372732214550405

Galinsky, A. D., Hugenberg, K., Groom, C., & Bodenhausen, G. V. (2003). The reappropriation of stigmatizing labels: Implications for social identity. In M. A. Neale & E. A. Mannix (Series Eds.) & J. T. Polzer (Vol. Ed.), *Research on managing groups and teams: Vol. 5. Identity issues in groups* (pp. 221–256). Emerald Publishing. https://doi.org/10.1016/S1534-0856(02)05009-0

Gapinski, K. D., Schwartz, M. B., & Brownell, K. D. (2006). Can television change anti-fat attitudes and behavior? *Journal of Applied Biobehavioral Research*, *11*(1), 1–28. https://doi.org/10.1111/j.1751-9861.2006.tb00017.x

Gast, A., Richter, J., & Ruszpel, B. (2020). Is there evidence for unaware evaluative conditioning in a valence contingency learning task? *Cognition and Emotion*, *34*(1), 57–73. https://doi.org/10.1080/02699931.2019.1652147

Gast, A., & Rothermund, K. (2010). When old and frail is not the same: Dissociating category and stimulus effects in four implicit attitude measurement methods. *Quarterly Journal of Experimental Psychology*, *63*(3), 479–498. https://doi.org/10.1080/17470210903049963

Gawronski, B. (2023). Automaticity and implicit measures. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (Vol. 3, pp. 404–426). Cambridge University Press.

Gawronski, B., & Bodenhausen, G. V. (2005). Accessibility effects on implicit social cognition: The role of knowledge activation and retrieval experiences. *Journal of Personality and Social Psychology*, *89*(5), 672–685. https://doi.org/10.1037/0022-3514.89.5.672

Gawronski, B., & Walther, E. (2012). What do memory data tell us about the role of contingency awareness in evaluative conditioning? *Journal of Experimental Social Psychology*, *48*(3), 617–623. https://doi.org/10.1016/j.jesp.2012.01.002

Gawronski, B., & Ye, Y. (2014). What drives priming effects in the affect misattribution procedure? *Personality and Social Psychology Bulletin*, *40*(1), 3–15. https://doi.org/10.1177/0146167213502548

Greenwald, A. G., & Lai, C. K. (2020). Implicit social cognition. *Annual Review of Psychology*, *71*(1), 419–445. https://doi.org/10.1146/annurev-psych-010419-050837

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (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

Greenwald, A. G., Nosek, B. A., Banaji, M. R., & Klauer, K. C. (2005). Validity of the salience asymmetry interpretation of the Implicit Association Test: Comment on Rothermund and Wentura (2004). *Journal of Experimental Psychology: General*, *134*(3), 420–425. https://doi.org/10.1037/0096-3445.134.3.420

Greenwald, A. G., & Pettigrew, T. F. (2014). With malice toward none and charity for some: Ingroup favoritism enables discrimination. *American Psychologist*, *69*(7), 669–684. https://doi.org/10.1037/a0036056

Hardin, J. (2015). Christianity, fat talk, and Samoan pastors: Rethinking the fat-positive-fat-stigma framework. *Fat Studies*, *4*(2), 178–196. https://doi.org/10.1080/21604851.2015.1015924

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*(3), 390–421. https://doi.org/10.1037/a0018916

Hunger, J. M., Smith, J. P., & Tomiyama, A. J. (2020). An evidence-based rationale for adopting weight-inclusive health policy. *Social Issues and Policy Review*, *14*(1), 73–107. https://doi.org/10.1111/sipr.12062

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, *25*(6), 881–919. https://doi.org/10.1111/j.1467-9221.2004.00402.x

Joy-Gaba, J. A., & Nosek, B. A. (2010). The surprisingly limited malleability of implicit racial evaluations. *Social Psychology*, *41*(3), 137–146. https://doi.org/10.1027/1864-9335/a000020

Kim, D. (2003). Voluntary controllability of the Implicit Association Test (IAT). *Social Psychology Quarterly*, *66*(1), 83–96. https://doi.org/10.2307/3090143

Kurdi, B., & Banaji, M. R. (2017). Repeated evaluative pairings and evaluative statements: How effectively do they shift implicit attitudes? *Journal of Experimental Psychology: General*, *146*(2), 194–213. https://doi.org/10.1037/xge0000239

Kurdi, B., & Banaji, M. R. (2019). Attitude change via repeated evaluative pairings versus evaluative statements: Shared and unique features. *Journal of Personality and Social Psychology*, *116*(5), 681–703. https://doi.org/10.1037/pspa0000151

Kurdi, B., & Charlesworth, T. E. S. (2023). A 3D framework of implicit attitude change. *Trends in Cognitive Sciences*, *27*(8), 745–758. https://doi.org/10.1016/j.tics.2023.05.009

Kurdi, B., & Dunham, Y. (2020). Propositional accounts of implicit evaluation: Taking stock and looking ahead. *Social Cognition*, *38*, s42–s67. https://doi.org/10.1521/soco.2020.38.supp.s42

Kurdi, B., & Mandelbaum, E. (2023). The case against implicit bias fatalism. *Nature Reviews Psychology*, *2*(11), 656–657. https://doi.org/10.1038/s44159-023-00248-y

Kurdi, B., Morehouse, K. N., & Dunham, Y. (2023). 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*, *124*(6), 1174–1202. https://doi.org/10.1037/pspa0000329

Kurdi, B., Sanchez, A., Dasgupta, N., & Banaji, M. R. (2024). (When) do counterattitudinal exemplars shift implicit racial evaluations? Replications and extensions of Dasgupta and Greenwald (2001). *Journal of Personality and Social Psychology*, *126*(4), 543–565. https://doi.org/10.1037/pspa0000370

Lai, C. K., & Banaji, M. R. (2020). The psychology of implicit intergroup bias and the prospect of change. In D. Allen & R. Somanathan (Eds.), *Difference without domination: Pursuing justice in diverse democracies* (pp. 115–146). University of Chicago Press. https://doi.org/10.7208/chicago/9780226681368.003.0005

Lai, C. K., Drake, R. G., & Beatty, J. F. (2023). Diversity training is just teaching. In E. Halperin, B. Hameiri, & R. Littman (Eds.), *Psychological intergroup interventions: Evidence-based approaches to improve intergroup relations* (pp. 58–68). Routledge. https://doi.org/10.4324/9781003288251-6

Lai, C. K., Hoffman, K. M., & Nosek, B. A. (2013). Reducing implicit prejudice. *Social and Personality Psychology Compass*, *7*(5), 315–330. https://doi.org/10.1111/spc3.12023

Lai, C. K., & Le Forestier, J. M. (2024, November 1). *Supplement: A comparative investigation of interventions to reduce anti-fat prejudice across five implicit measures*. https://osf.io/tdr2g

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J.-E. L., Joy-Gaba, J. A., Ho, A. K., Teachman, B. A., Wojcik, S. P., Koleva, S. P., Frazier, R. S., Heiphetz, L., Chen, E. E., Turner, R. N., Haidt, J., Kesebir, S., Hawkins, C. B., Schaefer, H. S., Rubichi, S., … Nosek, B. A. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*(4), 1765–1785. https://doi.org/10.1037/a0036260

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

Lai, C. K., & Wilson, M. E. (2021). Measuring implicit intergroup biases. *Social and Personality Psychology Compass*, *15*(1), Article e12573. https://doi.org/10.1111/spc3.12573

Lee, M., Ata, R. N., & Brannick, M. T. (2014). Malleability of weight-biased attitudes and beliefs: A meta-analysis of weight bias reduction interventions. *Body Image*, *11*(3), 251–259. https://doi.org/10.1016/j.bodyim.2014.03.003

Logel, C., Stinson, D. A., & Brochu, P. M. (2015). Weight loss is not the answer: A well-being solution to the "obesity problem". *Social and Personality Psychology Compass*, *9*(12), 678–695. https://doi.org/10.1111/spc3.12223

Mann, N. H., & Kawakami, K. (2012). The long, steep path to equality: Progressing on egalitarian goals. *Journal of Experimental Psychology: General*, *141*(1), 187–197. https://doi.org/10.1037/a0025602

Mann, T. C., Cone, J., Heggeseth, B., & Ferguson, M. J. (2019). Updating implicit impressions: New evidence on intentionality and the affect misattribution procedure. *Journal of Personality and Social Psychology*, *116*(3), 349–374. https://doi.org/10.1037/pspa0000146

March, D. S., Gaertner, L., & Olson, M. A. (2021). Danger or dislike: Distinguishing threat from negative valence as sources of automatic anti-Black bias. *Journal of Personality and Social Psychology*, *121*(5), 984–1004. https://doi.org/10.1037/pspa0000288

Marini, M., Rubichi, S., & Sartori, G. (2012). The role of self-involvement in shifting IAT effects. *Experimental Psychology*, *59*(6), 348–354. https://doi.org/10.1027/1618-3169/a000163

Marini, M., Sriram, N., Schnabel, K., Maliszewski, N., Devos, T., Ekehammar, B., Wiers, R., HuaJian, C., Somogyi, M., Shiomura, K., Schnall, S., Neto, F., Bar-Anan, Y., Vianello, M., Ayala, A., Dorantes, G., Park, J., Kesebir, S., Pereira, A., … Nosek, B. A. (2013). Overweight people have low levels of implicit weight bias, but overweight nations have high levels of implicit weight bias. *PLOS ONE*, *8*(12), Article e83543. https://doi.org/10.1371/journal.pone.0083543

Matharu, K., Shapiro, J. F., Hammer, R. R., Kravitz, R. L., Wilson, M. D., & Fitzgerald, F. T. (2014). Reducing obesity prejudice in medical education. *Education for Health*, *27*(3), 231–237. https://doi.org/10.4103/1357-6283.152176

Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology*, *65*(3), 469–485. https://doi.org/10.1037/0022-3514.65.3.469

Moskowitz, G. B., & Ignarri, C. (2009). Implicit volition and stereotype control. *European Review of Social Psychology*, *20*(1), 97–145. https://doi.org/10.1080/10463280902761896

Moskowitz, G. B., & Li, P. (2011). Egalitarian goals trigger stereotype inhibition: A proactive form of stereotype control. *Journal of Experimental Social Psychology*, *47*(1), 103–116. https://doi.org/10.1016/j.jesp.2010.08.014

Nosek, B. A., Bar-Anan, Y., Sriram, N., Axt, J., & Greenwald, A. G. (2014). Understanding and using the brief Implicit Association Test: Recommended scoring procedures. *PLOS ONE*, *9*(12), Article e110938. https://doi.org/10.1371/journal.pone.0110938

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*(1), 36–88. https://doi.org/10.1080/10463280701489053

O'Brien, K. S., Puhl, R. M., Latner, J. D., Mir, A. S., & Hunter, J. A. (2010). Reducing anti-fat prejudice in preservice health students: A randomized trial. *Obesity*, *18*(11), 2138–2144. https://doi.org/10.1038/oby.2010.79

Olson, M. A., Fazio, R. H., & Han, H. A. (2009). Conceptualizing personal and extrapersonal associations. *Social and Personality Psychology Compass*, *3*(2), 152–170. https://doi.org/10.1111/j.1751-9004.2008.00164.x

Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021). Prejudice reduction: Progress and challenges. *Annual Review of Psychology*, *72*(1), 533–560. https://doi.org/10.1146/annurev-psych-071620-030619

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*(3), 277–293. https://doi.org/10.1037/0022-3514.89.3.277

Perry, S. P., Murphy, M. C., & Dovidio, J. F. (2015). Modern prejudice: Subtle, but unconscious? The role of Bias Awareness in Whites' perceptions of personal and others' biases. *Journal of Experimental Social Psychology*, *61*, 64–78. https://doi.org/10.1016/j.jesp.2015.06.007

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123–205). Academic Press. https://doi.org/10.1016/S0065-2601(08)60214-2

Puhl, R. M., & Brownell, K. D. (2001). Bias, discrimination, and obesity. *Obesity Research*, *9*(12), 788–805. https://doi.org/10.1038/oby.2001.108

Puhl, R. M., & Heuer, C. A. (2009). The stigma of obesity: A review and update. *Obesity*, *17*(5), 941–964. https://doi.org/10.1038/oby.2008.636

Puhl, R. M., & Heuer, C. A. (2010). Obesity stigma: Important considerations for public health. *American Journal of Public Health*, *100*(6), 1019–1028. https://doi.org/10.2105/AJPH.2009.159491

Puhl, R. M., Moss-Racusin, C. A., Schwartz, M. B., & Brownell, K. D. (2008). Weight stigmatization and bias reduction: Perspectives of overweight and obese adults. *Health Education Research*, *23*(2), 347–358. https://doi.org/10.1093/her/cym052

Ranganath, K. A., Smith, C. T., & Nosek, B. A. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement methods. *Journal of Experimental Social Psychology*, *44*(2), 386–396. https://doi.org/10.1016/j.jesp.2006.12.008

Rooth, D.-O. (2009). Obesity, attractiveness, and differential treatment in hiring: A field experiment. *The Journal of Human Resources*, *44*(3), 710–735. https://doi.org/10.3368/jhr.44.3.710

Saguy, A. C., & Ward, A. (2011). Coming out as fat: Rethinking stigma. *Social Psychology Quarterly*, *74*(1), 53–75. https://doi.org/10.1177/0190272511398190

Schmidt, J. R., & De Houwer, J. (2012). Contingency learning with evaluative stimuli: Testing the generality of contingency learning in a performance paradigm. *Experimental Psychology*, *59*(4), 175–182. https://doi.org/10.1027/1618-3169/a000141

Schwarz, N. (2007). Attitude construction: Evaluation in context. *Social Cognition*, *25*(5), 638–656. https://doi.org/10.1521/soco.2007.25.5.638

Sim, M., Almaraz, S. M., & Hugenberg, K. (2022). Stereotyping at the intersection of race and weight: Diluted threat stereotyping of obese Black men. *Journal of Experimental Social Psychology*, *99*, Article 104274. https://doi.org/10.1016/j.jesp.2021.104274

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, *15*(1), 72–101. https://doi.org/10.2307/1412159

Sriram, N., & Greenwald, A. G. (2009). The Brief Implicit Association Test. *Experimental Psychology*, *56*(4), 283–294. https://doi.org/10.1027/1618-3169.56.4.283

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

SturtzSreetharan, C., Brewis, A., Hardin, J., Trainer, S., & Wutich, A. (2021). *Fat in four cultures: A global ethnography of weight*. University of Toronto Press.

Swift, J. A., Tischler, V., Markham, S., Gunning, I., Glazebrook, C., Beer, C., & Puhl, R. (2013). Are anti-stigma films a useful strategy for reducing weight bias among trainee healthcare professionals? Results of a pilot randomized control trial. *Obesity Facts*, *6*(1), 91–102. https://doi.org/10.1159/000348714

Talumaa, B., Brown, A., Batterham, R. L., & Kalea, A. Z. (2022). Effective strategies in ending weight stigma in healthcare. *Obesity Reviews*, *23*(10), Article e13494. https://doi.org/10.1111/obr.13494

Teachman, B. A., Gapinski, K. D., Brownell, K. D., Rawlins, M., & Jeyaram, S. (2003). Demonstrations of implicit anti-fat bias: The impact of providing causal information and evoking empathy. *Health Psychology*, *22*(1), 68–78. https://doi.org/10.1037/0278-6133.22.1.68

Teige-Mocigemba, S., & Klauer, K. C. (2008). "Automatic" evaluation? Strategic effects on affective priming. *Journal of Experimental Social Psychology*, *44*(5), 1414–1417. https://doi.org/10.1016/j.jesp.2008.04.004

Vartanian, L. R. (2010). Disgust and perceived control in attitudes toward obese people. *International Journal of Obesity*, *34*(8), 1302–1307. https://doi.org/10.1038/ijo.2010.45

Vartanian, L. R., Thomas, M. A., & Vanman, E. J. (2013). Disgust, contempt, and anger and the stereotypes of obese people. *Eating and Weight Disorders*, *18*(4), 377–382. https://doi.org/10.1007/s40519-013-0067-2

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*(1), 101–126. https://doi.org/10.1037/0033-295X.107.1.101

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, *72*(2), 262–274. https://doi.org/10.1037/0022-3514.72.2.262

Xu, K., Nosek, B., & Greenwald, A. G. (2014). Psychology data from the Race Implicit Association Test on the project implicit demo website. *Journal of Open Psychology Data*, *2*(1), Article e3. https://doi.org/10.5334/jopd.ac

Received January 19, 2024
Revision received November 1, 2024
Accepted November 17, 2024 ■