ANNUAL REVIEWS

Annual Review of Psychology

Prejudice Reduction: Progress and Challenges

Elizabeth Levy Paluck,[1] Roni Porat,[1,2] Chelsey S. Clark,[1] and Donald P. Green[3]

[1] Department of Psychology, Princeton University, Princeton, New Jersey 08540, USA; email: epaluck@princeton.edu

[2] Departments of Political Science and International Relations, Hebrew University, 91905 Jerusalem, Israel

[3] Department of Political Science, Columbia University, New York, NY 10027, USA

ANNUAL REVIEWS CONNECT

www.annualreviews.org

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

Keywords

anti-bias, anti-racism, meta-analysis, randomized control trials, interventions, diversity training

Abstract

The past decade has seen rapid growth in research that evaluates methods for reducing prejudice. This essay reviews 418 experiments reported in 309 manuscripts from 2007 to 2019 to assess which approaches work best and why. Our quantitative assessment uses meta-analysis to estimate average effects. Our qualitative assessment calls attention to landmark studies that are noteworthy for sustained interventions, imaginative measurement, and transparency. However, 76% of all studies evaluate light touch interventions, the long-term impact of which remains unclear. The modal intervention uses mentalizing as a salve for prejudice. Although these studies report optimistic conclusions, we identify troubling indications of publication bias that may exaggerate effects. Furthermore, landmark studies often find limited effects, which suggests the need for further theoretical innovation or synergies with other kinds of psychological or structural interventions. We conclude that much research effort is theoretically and empirically ill-suited to provide actionable, evidence-based recommendations for reducing prejudice.

Contents

INTRODUCTION

The pace of psychological research on prejudice reduction has accelerated since our 2009 review (Paluck & Green 2009), which found a diverse array of theoretical approaches but a dearth of studies that convincingly demonstrate the effectiveness of prejudice reduction strategies. Prejudice reduction is now recognized as a subfield of its own, inspired by and yet distinct from the large literature documenting and describing prejudice. The rapid growth of this subfield reflects the contributions of many social sciences in addition to psychology, and it features a broadening of the research agenda to new and geographically diverse contexts.

In this article, we cover the burgeoning literature from the past decade. Our review encompasses both studies of the basic science of prejudice reduction and evaluations of fully developed prejudice reduction intervention programs, some of which have been tested in real-world settings. Recognizing the increased emphasis that behavioral scientists and policy makers place on credible causal inference, we restrict our review to experimental studies. Using replicable search criteria, we assembled 418 experimental studies reported in 309 manuscripts. We summarize and critically evaluate these recent contributions to the literature both quantitatively and qualitatively. Our quantitative review uses meta-analysis to assess the extent to which interventions inspired by different theoretical approaches reduce prejudice. For example, how successful is cognitive or emotional training, contact, or social categorization? We consider intervention success with different targets of prejudice (e.g., groups defined by ethnicity, religion, sexuality, etc.) and ask to what extent success depends on different ways of gauging prejudice (e.g., explicit versus implicit

Prejudice: a negative bias toward a social category of people, with cognitive, affective, and behavioral components

attitudes) or on the time frame within which measurement occurs (e.g., immediately after exposure versus days or weeks later). Our study sample reveals recent trends in prejudice reduction research, such as the study of minor or light touch interventions and of synergistic theoretical tests. Our quantitative synopsis of the literature also uncovers troubling methodological issues. Principally, we find a pattern of smaller studies reporting significantly stronger effects, a symptom of publication bias. One urgent message of this article is the need for caution when reviewing prejudice reduction recommendations generated by small studies with optimistic conclusions.

We complement this quantitative assessment with a descriptive review of the studies and their characteristics. When we chart trends in theoretical approaches, we observe a new focus on mentalizing as a salve for prejudice—interventions aimed at changing an individual's cognition, perception, or even imagination, rather than interpersonal communication or collaboration. Methodologically, we describe what happens when scientists take theories developed in controlled laboratory settings and experimentally test them in the settings to which the laboratory studies hope to generalize. We identify a group of landmark studies from the past decade that are large, well-crafted experiments in real-world settings testing fundamental psychological questions of prejudice reduction. These studies address the effects of interpersonal influence, intergroup contact, cooperation, exchange of perspectives, and narratives. Their findings are convincing but also sobering. The fact that landmark studies often find limited effects on prejudice suggests the need for further theoretical innovation or synergies with other kinds of psychological or structural interventions.

We structure our review as follows: First, we specify our search and selection methodology for assembling our collection of recent studies. After describing basic trends in the pace and theoretical focus of research during the past decade, we provide meta-analytic estimates of the efficacy of prejudice reduction efforts inspired by various theoretical approaches, noting along the way gaps in what we know or anomalies that require further investigation. The last section of the article reviews important recent studies that advance both what we know about prejudice reduction and how to study it.

We conclude that much research effort is theoretically and empirically misguided if the aim is to provide actionable, robust, evidence-based recommendations for reducing prejudice in the world. Our final section offers suggestions about how the literature might be refocused so that it can speak to the urgent need for workable and proven prejudice reduction interventions.

TRENDS IN THEORY AND INTERVENTION IN THE PAST DECADE

What types of interventions have received the most research attention from social science over the past decade? Modal studies—which account for one-third of all prejudice reduction research—test an intervention that involves second-hand or imagined contact with outgroups. The scholarly emphasis on contact that occurs vicariously or via an imagined interaction is a relatively recent development; our previous review (Paluck & Green 2009) found relatively few studies of this kind. The next largest group of studies involves some kind of cognitive or emotional training. These interventions, which train individuals to use thinking and emotion regulation strategies to fight off their personal prejudices, are tested in more than one-quarter of recent scholarly work. The emphasis on this type of work stands in contrast to interventions that attempt to provide information that breaks down negative stereotypes or promotes cross-cultural understanding, which have become relatively rare. Social categorization interventions, which encourage participants to rethink group boundaries or to prioritize common identities shared with specific outgroups, constitute the third largest group of studies. Another moderately popular strategy with a long intellectual pedigree involves reminding individuals of their own or their group's values or past egalitarian

Light touch interventions: treatments that are easy to implement, brief (under 10 minutes), inexpensive, and thought to have lasting effects

Experiment: a study that assigns participants to treatment and control conditions using some random procedure, such as a coin flip

behaviors, leveraging preferences for self- and group-based consistency to urge individuals toward an anti-prejudice stance. A small share of studies involves some kind of direct contact with outgroups or, relatedly, peer discussion or dialogue with or about them.

Noteworthy is the scarcity of prejudice reduction research on the kinds of programs that are frequently called for in the real world, specifically entertainment and mass media interventions, as well as diversity training and sensitivity or cultural competence training for leaders, care providers, and decision makers in business, health services, and law and policing. Rigorous scholarly research does not always reflect what is popularly demanded, although some of the basic research purports to test methods that could be used in training curricula.

The burgeoning research literature offers a number of theoretical puzzles. For example, the surge in popularity of light touch interventions exposes a gap in understanding about the effects of intervention dosage and intensity. After a decade of experimental research on both brief low-cost interventions and more intensive long-term interventions, we now face the question: Do we observe larger or longer-lasting changes from the more intensive interventions? If not, what is responsible for the large effects of brief treatments? To this question, we note that some of the larger and more time-intensive treatments tested in the world by landmark studies show remarkably modest effects. We find large effects of light touch interventions tested in the laboratory, but our meta-analysis shows that these large effects tend to coincide with small N studies, which raises suspicions about publication bias for the reasons explained below.

Another theoretically intriguing development is that a growing number of prejudice reduction interventions tested in the field find that intergroup behaviors change without attitude change (Mousa 2020, Paluck 2009a, Scacco & Warren 2018). What is the implication for how we think about prejudiced attitudes and their correspondence to prejudiced behaviors? Is it plausible that attitudes and behaviors remain out of sync? Or do we lack the type of long-term measurement that shows behavior reverting back to be consistent with prior attitudes?

META-ANALYTIC SEARCH AND ANALYSIS

We followed biomedical meta-analytic standards when assembling the literature (Moher et al. 2009). We searched five separate databases for published and unpublished articles starting in 2007, the final year covered by our previous review (Paluck & Green 2009). We supplemented these searches with a text-based search in the proprietary database Articles+. We used 19 different keywords to describe prejudice and its various manifestations (e.g., "stereotyp*," "islamophobi*"), along with qualifying terms indicating an experimental test of an intervention, such as "experiment*" and "control group." For detailed information on our search, the reader is referred to the **Supplemental Appendix**.

Theoretical and Methodological Criteria for Study Selection

Our exhaustive search yielded over 16,000 manuscripts. The authors decided whether a study met eligibility criteria for inclusion. The PRISMA flowchart in the **Supplemental Appendix** explains how we winnowed the original pool of studies down to the 1,835 manuscripts that were reviewed in full.

Theoretically, we used a definition of prejudice that would encompass standard usage by researchers working in this domain: Prejudice is animus, or negative bias, toward social groups and their putative members. To be selected, the studies also needed to research an intervention that in some way sought to reduce prejudice as a psychological predisposition or its expression in behavior or behavioral intentions.

COMMON PROBLEMS IN EVALUATING PREJUDICE REDUCTION EFFORTS

Our review of prejudice reduction studies repeatedly encountered three methodological problems. The first one is attrition: Participants are lost to follow-up, fail to complete end-line surveys, or are excluded for failing to pay attention. Attrition jeopardizes the unbiasedness of a randomized experiment, particularly if attrition operates differently among treatment and control groups. In some studies a large share of participants were missing at end-line; in others, the rates of missing participants appeared significantly different across experimental conditions.

Second, in cluster-randomized experiments, groups of people, rather than individuals, are randomly assigned to experimental conditions. Reliably estimating standard errors requires at least 10 clusters, but many intervention studies do not approach even this minimal number. A further complication is that conventional estimators such as regression yield biased results when experiments use small numbers of clusters of unequal size.

Finally, we encountered widespread lack of transparency insofar as few studies made their data and code publicly available. Many studies neglected to report key statistics such as standard errors or selectively reported estimated treatment effects only for subsets of the data, and the lack of public data made it impossible for us to calculate the relevant statistics ourselves.

We excluded several related literatures, such as those studying the effects of prejudice on its targets (e.g., stereotype threat, social belonging, and self-affirmation). We excluded priming interventions that did not aim to reduce prejudice in an enduring way but rather aimed to demonstrate that prejudice is momentarily malleable. We also excluded work on intergroup bias that is not necessarily rooted in animus. For example, we did not include prejudice toward political groups, which is associated with a distinct theoretical literature, nor did we include work on interuniversity rivalries. As in the previous review (Paluck & Green 2009), this review does not include gender-based prejudice, with the exception of prejudice toward transgender or genderqueer identities. An additional criterion for study inclusion was the measurement of prejudice as an outcome. Thus, we did not include studies whose outcomes focused solely on policy preferences.

Methodologically, we excluded nonexperimental studies from our meta-analysis.[1] To be included, a study had to clearly indicate that random assignment was used to allocate participants or groups to experimental conditions (see the sidebar titled Common Problems in Evaluating Prejudice Reduction Efforts).

Coding and Reliability

In total we recovered 309 manuscripts from our search that qualified methodologically and theoretically based on our criteria. These papers were initially read by one of the authors, who coded high-level attributes, i.e., the type of intervention approach based on the theory tested (e.g., a contact intervention versus a peer influence intervention). The authors then identified the relevant experimental conditions for which statistics needed to be coded (see the **Supplemental Appendix** for details) and selected up to five prejudice-related outcomes for each study. We always coded a behavioral and an implicit outcome, if reported, as well as up to three additional outcomes, including explicit attitudes and beliefs, emotions, norms, and behavioral intentions. If studies reported additional outcomes, we took the outcomes assigned most priority in the abstract or paper by the

Supplemental Material >

---

[1]Meta-analysis presupposes that the reported standard errors properly summarize the statistical uncertainty associated with a given study. Because nonstatistical sources of error (e.g., confounding due to omitted variables) are not typically accounted for in observational studies, guides to best practices warn against intermingling experimental and nonexperimental studies.

authors, a rule that typically favors selecting positive results of the intervention. Finally, the authors coded whether the intervention might be considered to be light touch or not and whether the theoretical grounding of the intervention fused the insights of more than one theory to create a synergy or interaction.

After this phase, information from each study was recorded and then checked by at least two members of a team of nine master- or postdoctoral-level paid research assistants blind to the hypotheses of the study. This team coded both qualitative and quantitative attributes of the study, including all statistics needed to calculate the effect size of the outcomes, information about the intervention (e.g., time duration), the target population (e.g., age and context like school or work), the type of prejudice targeted (e.g., race, nationality, immigration status), the timing of the outcome measurement (e.g., immediately after the intervention or delayed), and whether the study followed open science practices (e.g., preregistration, open data). Coding disagreements were all resolved by documented discussions among the authors, and all codes were double checked by either a research assistant or an author.

Meta-Analytic Procedures

Methodological approaches to meta-analysis vary, and we settled on an approach that is robust to alternative specifications (see **Supplemental Table 1**). To distill the studies down to a single estimate, we averaged their Cohen's d's for each outcome and averaged their estimated sampling variance. These distilled estimates (one per study) were then used as inputs to a random effects meta-analysis. To address the issue of nonindependence across studies from the same article, we clustered the standard error of this random effects estimate by article. Our data and replication code may be found at Dataverse.[2]

A concern for any meta-analysis is publication bias, or the tendency for journals to favor manuscripts that report statistically significant results (see the sidebar titled Publication Bias in the Prejudice Reduction Literature). The preference for significant findings creates a "file drawer problem" (Rosenthal 1979) whereby studies with small effects go unpublished and unnoticed

PUBLICATION BIAS IN THE PREJUDICE REDUCTION LITERATURE

Publication bias occurs when the direction or strength of a study's outcome influences whether it is published or not. When academic journals are reluctant to publish research papers that report statistically insignificant treatment effects, studies that produce weak or null effects may remain invisible to the academic community.

A telltale sign of publication bias is a strong positive relationship between reported effects and their standard errors, because smaller studies, which tend to generate larger standard errors, must produce larger effect estimates in order to achieve significance at the 0.05 level. Our collection of studies displays a powerful relationship of this kind, even when we focus solely on lab experiments ($N = 301$): Lab studies that generate precise results tend to show less prejudice reduction than smaller studies that typically generate results with large standard errors. A linear regression of all effect sizes on standard error shows a distressing positive relationship (see the **Supplemental Appendix** for a graphical depiction) in which the intercept is close to zero, suggesting that a study large enough to generate a standard error of approximately zero would, on average, produce no change in prejudice at all. In other words, if the current collection of studies had been conducted on a much larger scale, our analysis would have shown no reduction in prejudice.

[2]The persistent Dataverse link can be found at **https://dataverse.harvard.edu/dataset.xhtml?persistentId= doi:10.7910/DVN/ODACR5**.

by those reviewing the literature; it also creates a "*p*-hacking" problem (Simonsohn et al. 2014) whereby researchers make statistical choices that militate in favor of producing larger, statistically significant estimates. As explained below, in every theoretical domain we find unmistakable indications of publication bias: Large-*N* lab, online, or field studies that generate precise results tend to produce much weaker effects than small-*N* studies that generate results with large standard errors. For this reason, our description of the literature is careful to present two sets of meta-analytic results: (*a*) a synoptic analysis that includes all studies and ignores publication bias, and (*b*) an analysis limited to studies with reasonably good statistical precision, i.e., those experiments that are in the highest quintile of number of participants in the intervention arm (here, 78 or more participants).

Overview of Studies in the Meta-Analytic Database

We begin by describing the database of prejudice reduction research. The database spans the years 2007–2019, encompassing 309 manuscripts that describe 418 studies, which report 1,292 distinct point estimates. This trove of recent research encompasses a wide array of theoretical ideas about how prejudice is best overcome. **Figure 1** summarizes the intervention approaches according to how we classified each point estimate or study. We describe each intervention approach below as we lay out the results for each category.

Figure 1 also offers a glimpse of the types of outgroup prejudices that have attracted recent scholarly interest. The modal category is racial and ethnic prejudice, which accounts for nearly half of all studies. This category would encompass almost 60% of all studies were it expanded to include nationality and religion, which are sometimes correlated with ethnicity. Migrants and refugees have also attracted recent scholarly attention. The other half of the literature is apportioned across a wide range of different targets of prejudice. Roughly 20% of all studies focus on physical attributes related to ability, age, and body size. Approximately 13% of studies focus on sexuality and gender expression. Relatively few studies nowadays generate fictive (minimal) groups to study prejudice reduction.

Whose prejudices have been the object of recent study? Nearly two-thirds of all studies attempt to reduce the prejudices of college students. Another 10% focus on students in grade school or high school. Almost one in five deploy interventions aimed at online respondents, including convenience samples of workers earning wages for completing surveys. The relatively small remainder of the studies analyzed underscores how rarely researchers study adults outside universities: members of the community (5.5%), people in the workplace (2.2%), or religious congregants (0.5%).

Figure 1 further characterizes the types of outcome measures used in the studies. Two-thirds of all study outcomes are some kind of explicit survey measure of attitudes or beliefs. Measures of implicit attitudes and behavioral intentions are much less common, at less than 10% each, with a smattering of studies reporting survey measures of emotion and empathy. Only 7% of outcomes are behavioral.

How have the volume of prejudice reduction research and the setting of the research evolved over time? Prior to 2010, fewer than 16 studies were generated annually, but by 2018 and 2019, the annual production was 58. Some of this increase reflects the advent of online experimental studies from 2014 on. Lab studies also contribute to the overall rate of growth. Although some growth is evident since 2012, field experiments never account for more than 4 studies in any given year (see **Supplemental Figure A**).

Supplemental Material >

META-ANALYTIC RESULTS

Our random effects meta-analysis of all prejudice-reduction interventions from 2007 to 2019 reveals an effect size *d* of 0.357, with a standard error (SE) of 0.02. To get a sense of what this estimate
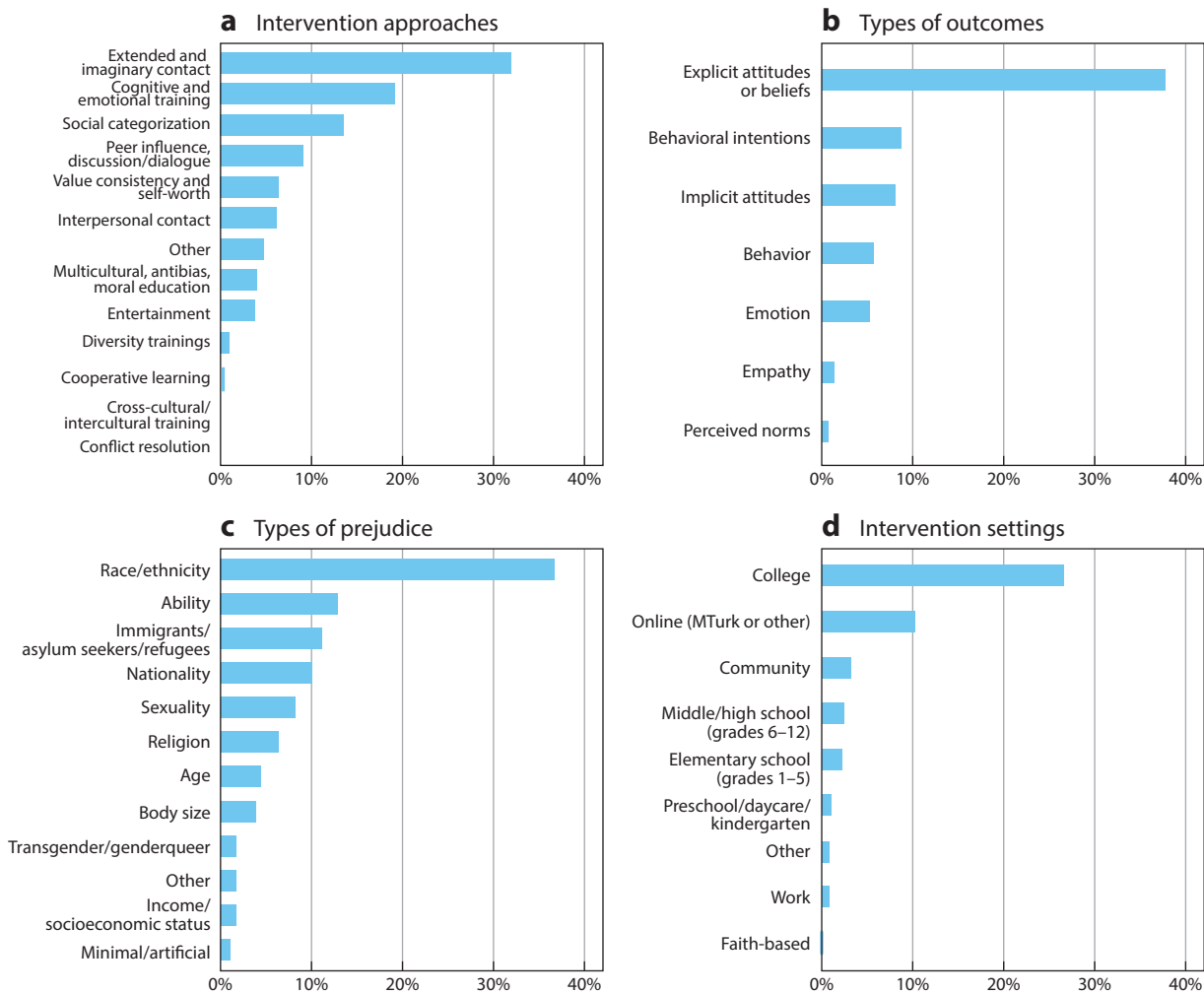
Figure 1

Intervention approach, types of outcomes measured, types of prejudices addressed, and intervention settings. In these figures, if a single study tests more than one intervention approach or more than one prejudice, it is counted in both codes (i.e., the denominator for all percentages adds up to more than the total number of studies in the meta-analysis, $N = 418$). For example, 35% of all studies report more than one type of outcome.

implies, consider responses to a feeling thermometer in which participants rate how warmly they feel toward another group on a 0 (cold) to 100 (warm) scale. For example, individuals who assigned on average a rating of 40 toward Black people, which indicates a mildly negative feeling, would on average be moved approximately 8 points to a rating of 48, nearly reaching a neutral feeling of 50 [using the standard deviation of 21.2 reported by the 2016 wave of the American National Election Survey (ANES) for feeling thermometers toward Black people]. To obtain $d = 0.357$, we averaged estimates from each study before meta-analyzing the study-level estimates and clustered the standard errors by paper. This estimate is highly robust to different estimation strategies (see the **Supplemental Appendix**), and we were able to reject a null of no treatment effect at $p < 0.0001$.

Table 1 All studies by quintile group

Treatment sample size	Number of studies	Number of articles	Effect size	Standard error
≤ 25	74	62	0.606	0.052
26–34	74	63	0.421	0.058
35–48	75	67	0.371	0.042
49–77	70	65	0.341	0.040
≥ 78	73	61	0.187	0.022

Although our result is robust to the precise manner in which meta-analytic results are calculated, the results are not robust to the most basic assessments of study quality. We offer just one: study size, or number of participants. In the absence of publication bias, we should obtain similar average effect estimates from small and from large studies. However, **Table 1** demonstrates a powerful inverse relationship between study size and effect size. Restricting attention solely to the quintile of smallest studies (i.e., the 74 studies that allocate 25 or fewer participants to the treatment condition), we obtain a meta-analytic estimate of $d = 0.61$ (SE = 0.05). This large effect size would on average move a person who feels mildly negatively toward Black people at 40 to a solidly neutral feeling of 53. By contrast, the 73 studies in the highest quintile of study population size, which allocate 78 or more participants to the treatment group, generate a meta-analytic estimate of $d = 0.19$ (SE = 0.02). These larger studies predict that on average, interventions would change feelings toward Black people in a positive direction but only by approximately 4 points on the scale, such that people who started out feeling 40 would still rate their feelings as mildly negative (44) following an intervention. Studies with intermediate-size treatment groups produce intermediate-size effects. The relationship between our meta-analytic effect size and the size of a study's treatment group is highly significant ($p < 0.0001$). Importantly, this finding is not an artifact of research methodology—e.g., of online studies with larger samples finding weak effects (see **Supplemental Table 2** and **Supplemental Figures 1–4**).

The size of the three average effects we have described—from small studies, large studies, and all studies combined—may be grasped intuitively by comparing it to the size of the well-known reduction in prejudice toward gay people in the United States over the past three decades. From 1984 to 2016, the ANES recorded that average feeling thermometer responses in relation to gay people went from 30.9 (cold) to 60.7 (warm), which is a d of 1.09. Thus, the average effect size of the prejudice reduction literature, which is an optimistic calculation ignoring publication bias, is approximately one-third of the shift in feelings toward gay people. A Cohen's d of 0.357 is certainly not transformative, but a collection of interventions that on average produce effects of 0.357 is certainly meaningful. The problem is, can we believe that the average effect is 0.357? The evidence reviewed above suggests that we cannot. Thus, as we review each type of intervention below, we continually revisit the question of whether the average effect of a particular type of intervention is corroborated by large studies.

WHICH PREJUDICE REDUCTION INTERVENTIONS WORK?

Applied Interventions

First, we review experimental studies that tested the effects of a multifaceted applied intervention, such as diversity training, antibias and moral education, and the like. These types of interventions typically involve more than one theoretical mechanism, and so experiments testing their outcomes are more akin to program evaluations than to theoretical tests. Evaluating combinations of theoretically driven activities can still be informative for theory, however, as we discuss below.

Antibias, multicultural, and moral education. These interventions come primarily from the tradition of prejudice reduction in the field of educational psychology. Antibias education and multicultural education draw variously on theories addressing the socialization of prejudice, cognitive and moral development, and learning. The form of these interventions also ranges widely. For example, Li et al. (2019) taught health care providers in China strategies for combating mental illness stigma, and Neto et al. (2016) taught to Portuguese sixth-graders a 6-month music program on anti–dark skin bias featuring an ethnically diverse pool of artists.

We found 20 experiments testing antibias, multicultural, and moral education programs across 18 papers. On average this group of approaches yielded an effect size d of 0.30 (SE = 0.06). However, when we restrict our sample to studies with a treatment N of 78 or more, we find only 5, and the effect size drops from 0.30 to 0.23 (SE = 0.08). These interventions for the most part measured explicit attitudes as outcomes (17 studies, which yielded an average d of 0.28, SE = 0.07); just 4 studies measured behavioral outcomes. Despite the fact that these interventions are meant to address real-world settings, only 5 of the experiments were conducted in actual educational or workplace settings; 12 were conducted in laboratory settings, and another 3 were tested with online experiments.

Two related instructional programs that featured prominently in our previous review (Paluck & Green 2009), cross-cultural and intercultural training and cooperative learning, were rarely studied in the past decade and numbered too few to be the object of a separate meta-analysis. Cross-cultural and intercultural training programs are most frequently implemented in workplaces, with managers and workers assigned to posts abroad in unfamiliar cultures. Cooperative learning was one of the most popular educational interventions in the 1980s and 1990s in the United States and was hypothesized to reduce student prejudice. It seems that experimental research on cross-cultural and intercultural training programs has slowed (Kraimer et al. 2016); the same may be said of cooperative learning, which has also fallen out of practice in the educational sector.

Diversity training. The notion of diversity training encompasses a wide category of interventions that are "designed to attack bias" among managers and workers (Kalev et al. 2006, p. 591). More generally, diversity trainings take place both in and outside of the workplace, using "a distinct set of instructional programs aimed at facilitating positive intergroup interactions, reducing prejudice and discrimination, and enhancing the skills, knowledge, and motivation of participants to interact with diverse others" (Bezrukova et al. 2016, p. 1228). Under this definition, previous reviews describing the positive effects of diversity training have included a startlingly wide range of interventions. For example, Bezrukova et al. (2016) included in their diversity training meta-analysis studies such as the one by Paluck (2009), which exposed Rwandans to radio soap operas on the theme of ethnic reconciliation. By contrast, in this review we categorize interventions as diversity training if they self-identified as a diversity training or as a related program, like sensitivity or cultural competence training. The latter programs are often used with health and legal services providers like nurses, doctors, and judges. For example, we include Smith & Silk (2011), who tested the effects of an interactive online simulation of a doctor-patient interaction on the level of comfort of medical students interacting with Muslim patients.

Several kinds of theories have inspired diversity training curricula. Some curricula rely on contact theory, and they bring diverse trainers or diverse members of an organization into communication with others in the organization (Paluck 2006). Other approaches rely on theories of cognitive and value-based consistency to motivate personal change by offering presumably well-intentioned trainees personalized information about their bias (e.g., Perry 2011); still others

rely on behavioral insights about the efficacy of providing concrete strategies by which trainees can debias their decisions and behaviors (Chang et al. 2019).

After including studies in this category only if they self-identified as interventions of diversity, sensitivity, or cultural competence training, we count just six experimental studies reported in six manuscripts during the last decade. The average meta-analytic effect size of these few studies is a d of 0.3 (SE = 0.16). However, if we restrict the sample of diversity training studies to those in the quintile with the largest participant samples, only two studies remain. The average effect of those studies, by Chang et al. (2019) and Smith & Silk (2011), is substantially lower ($d = 0.07$). Notably, Chang et al. (2019) is also the only experiment testing a diversity training as it would be naturally implemented in the world, in collaboration with a large corporate employer. Four of the studies coded as diversity training were actually conducted in a university setting, and these are the studies that drive up the average effect size—they are associated with a d of 0.45 (SE = 0.23).

Overall, contrary to a previous review (Bezrukova et al. 2016), we did not find a broad evidence base on which to draw conclusions about the efficacy of diversity training. This is disappointing, considering the frequency with which calls for diversity training emerge in the wake of widely publicized instances of discriminatory conduct. While the small number of experimental studies provide encouraging average effects, details of these studies reveal that the effects shrink when the trainings are conducted in real-world workplace settings, when the outcomes are measured at a greater time distance than immediately following the intervention, and, most importantly, when the sample size is large enough to produce reliable results.

Basic Research Interventions

Next, we review interventions explicitly designed to test theoretical perspectives. Experiments testing these interventions are considered part of the basic research approach to prejudice reduction because the experimental manipulation seeks to isolate theoretical mechanisms.

Cognitive and emotional training. Interventions categorized as cognitive and emotional training share the idea that individuals can be trained to use thinking and emotion regulation strategies to fight off their personal implicit or explicit prejudices. These forms of training break down into three loose types. The first type is best characterized as cognitive conditioning, and it seeks to alter a particular aspect of a person's cognitive association with or assessment of an outgroup through practice or repeated contradictory pairings (Devine & Monteith 1999, Fiske & Lin 81, Olson & Fazio 2006). For the most part, this training targets automatic or implicit negative associations (e.g., Muslim = dangerous). For example, interventions seeking to automate more positive responses to outgroup members may train people to have an "approach" response to Black faces (e.g., by pulling a joystick toward themselves when Black faces appear on a screen; see Kawakami et al. 2007), to make "if-then" plans when encountering a stigmatized group member (e.g., instructions to think "safe" thoughts when encountering Black people; see Mendoza et al. 2010, Stewart & Payne 2008), and to consider situational rather than dispositional factors when evaluating an outgroup member (Latu 2010, Walsh 2013). Other interventions use classical conditioning techniques, such as pairing representations of members of stigmatized groups with positive stimuli (e.g., Balas & Sweklej 2013, French et al. 2013, Woodcock & Monteith 2013).

A fundamental assumption shared by the second type of intervention is that emotions contribute to the expression of prejudice. Cognitive-emotional interventions attempt to change emotional experiences through various direct or indirect cognitive techniques. For example, one

intervention trained people in emotion regulation (e.g., Gross 2013) and, specifically, to reduce their negative emotions when encountering situations related to the outgroup (e.g., Halperin et al. 2013). Another type of intervention utilizes guided meditation techniques to accentuate positive emotions toward others like gratitude and love. For example, in one study, participants received a 7-minute guided loving kindness meditation in which they were asked to close their eyes, relax, take deep breaths, and imagine people who deeply cared for them sending them love; they were then instructed to open their eyes and send this love back toward a photo of a Black person (Stell & Farsides 2016). Another intervention combines both cognitive and emotional strategies to combat implicit bias. The prejudice habit-breaking intervention (e.g., Devine et al. 2012, Forscher et al. 2017) focuses on changing implicit bias by raising awareness of one's own biases and providing strategies to overcome them.

A third type of cognitive and emotional prejudice reduction intervention uses perspective taking, a technique whereby individuals envision themselves in another person's shoes. Theory predicts that by encouraging individuals to experience the thoughts or emotions of another person, perspective taking increases the self/other overlap and leads to more favorable attitudes toward that particular person but also toward members of the person's group (Galinsky & Moskowitz 2000). Canonical perspective-taking interventions give direct instructions; for example, they might ask participants to write a short essay from the point of view of an outgroup member (e.g., Berthold et al. 2013, Castillo et al. 2014). Other interventions go to greater lengths, such as the one by Berthold et al. (2013), who asked participants to take the perspective of the elderly by climbing up a staircase while wearing restrictive knee pads that made it harder for them to move (Berthold et al. 2013). Recent interventions make use of technology to create a more immersive perspective-taking experience—for example, by using virtual reality to make individuals see themselves as another person (e.g., Banakou et al. 2016, Oh et al. 2016).

Overall, various types of cognitive and emotional training were the object of 107 studies, discussed in 75 manuscripts, over the past decade. As for their apparent effects on prejudice reduction, meta-analysis reveals a substantial average effect size d of 0.35 (SE = 0.05). However, when examining the studies with a treatment N of 78 or greater, we are left with just 25 studies that reveal an average effect size d of 0.22 (SE = 0.05). Thus, publication bias seems to be a concern when reading research in this area.

The training interventions in the full sample of cognitive and emotional training studies addressed a broad array of prejudices but focused primarily on race and ethnic prejudice ($d = 0.35$, SE = 0.06, $N = 63$) and on explicit attitudes and beliefs ($d = 0.26$, SE = 0.05, $N = 64$). The vast majority of the studies involved college students ($d = 0.37$, SE = 0.06, $N = 72$), took place in the laboratory ($d = 0.39$, SE = 0.06, $N = 79$), and measured prejudice immediately after the intervention's conclusion ($d = 0.33$, SE = 0.05, $N = 98$).

Value consistency and self-worth. Value consistency and self-worth interventions draw on what is believed to be a universal human desire to maintain a consistent and positive image of the self and of the group (Festinger 1962, Tajfel & Turner 1979). These interventions include reminders of individuals' or their group's egalitarian preferences or history in order to inspire consistency with that history in the present moment (e.g., Heitland & Bohner 2010, Smeekes et al. 2012), remind people of moral exemplars (Witkowska et al. 2019), and provoke introspection about one's existing beliefs and prejudices (Radnitz 2018). Affirmation experiments, which ask individuals to recall their or their group's important values (e.g., Sherman & Cohen 2006), predict that an individual given the opportunity to feel affirmed by their values will be less likely to express prejudice (e.g., Lehmiller et al. 2010, Villicana et al. 2018).

Also included in this category is an intervention that has attracted increasing scholarly interest: the prejudice confrontation intervention, in which individuals are given feedback about their level of prejudice. This intervention presumes that most people are motivated to not be prejudiced; upon receiving feedback that they have high levels of unconscious prejudice, they are motivated to bring their cognition and behavior in line with their aspirational self-image (e.g., Chaney 2016, Perry 2011). Note that this type of intervention may be difficult to administer to members of the general population, who may push back against the notion that they harbor prejudice or may be less motivated to be unprejudiced compared to typical laboratory study participants. In this regard, the mostly lab-based literature speaks to the theoretical efficacy of prejudice confrontation, absent the complications of orchestrating real-world confrontations.

Our meta-analysis reveals 35 experiments across 22 manuscripts that can speak to values and self-consistency. The average effect size d is 0.41 (SE = 0.09). These interventions have never been tested in the field; in the lab the effect size is higher ($d = 0.50$, SE = 0.14, $N = 20$) than it is online ($d = 0.30$, SE = 0.10, $N = 15$). The strongest effects apply to changes in explicit attitudes ($d = 0.45$, SE = 0.09, $N = 30$) and emotional responses ($d = 0.42$, SE = 0.20, $N = 4$), as opposed to behavioral intentions ($d = 0.06$, SE = 0.12, $N = 5$). Again, when we restrict the sample to studies with a treatment N greater than or equal to 78, the number of studies drops to 5, and the average effect size d becomes 0.29 (SE = 0.17).

Peer influence, discussion, and dialogue. This category of interventions is united by the idea that people who share important identities, peers, or ingroup members have a powerful influence over one another's impression of the attitudes and behaviors that are typical, desirable, and correct (Goldstein et al. 2008, Prentice & Paluck 2020). The interventions in this category wield peer influence in various ways to reduce prejudice. Some interventions use peers as messengers for a cause, such as evangelical leaders messaging to other conservative Christian activists about tolerance toward immigrants or LGBT individuals (Robinson 2010). Others, sometimes labeled social pressure or social consensus interventions, leverage summaries of peer norms in which individuals are told about the more tolerant beliefs of their peers and are motivated to move their own attitudes toward the peer group (e.g., Gómez et al. 2018, Patel 2013, Sechrist & Milford-Szafran 2011). Finally, a tradition of bringing together individuals in dialogue leverages both peer influence and other psychological processes such as group interaction and information transmission to make intergroup dialogue an effective prejudice reduction intervention (see Gurin et al. 2013). In the last decade, experimenters have tested interventions that use dialogue to increase White students' desire and ability to be allies to other racial and ethnic groups (Alimo 2012) and to encourage secondary teachers to build a more inclusive and safe space for lesbian, gay, and bisexual students (Dessel 2010).

Across the 40 peer influence studies reported in 31 articles over the past decade, the average prejudice reduction effect of dialogue interventions was slightly weaker compared to other approaches, with a d of 0.27 (SE = 0.07). Further restricting the sample to the 10 studies whose treatment N is equal to or greater than 78, we find a smaller average effect ($d = 0.2$, SE = 0.05).

The prejudices addressed by the collection of peer influence interventions reviewed here were varied. The prejudices most successfully addressed were those toward immigrants, asylum seekers, and refugees ($d = 0.34$, SE = 0.11, $N = 10$) and toward LGBT individuals ($d = 0.39$, SE = 0.11, $N = 11$). Eleven studies measured behavior ($d = 0.63$, SE = 0.20, $N = 5$) or behavioral intentions ($d = 0.38$, SE = 0.03, $N = 6$) as an outcome, which is higher than other areas of prejudice reduction research. Another promising result is that the effect of peer influence measured at least one day after the intervention was on average stronger than the one measured immediately following the intervention ($d = 0.31$, SE = 0.08 versus $d = 0.26$, SE = 0.08). However, as with all interventions,

the number of studies measuring effects one day or more after the intervention was small compared to the number of studies providing an immediate measurement (7 versus 34 studies).

Social categorization. Social categorization interventions are based on a suite of theories arguing that the simple act of categorizing others into ingroups and outgroups is enough to foster intergroup bias and, conversely, to reduce it when group boundaries are rearranged or questioned (e.g., Gaertner & Dovidio 2000, Park & Rothbart 1982, Tajfel & Turner 1979).

Methods of intervening on this basic phenomenon fall into two major categories: modifying intergroup boundaries and changing the perceptions of groups within the existing boundaries. To modify intergroup boundaries, interventions use various approaches to make the outgroup seem more similar to the ingroup or even to demonstrate that the two groups share a common identity (Gaertner & Dovidio 2000). For example, Hall et al. (2009) asked participants to list characteristics shared by the ingroup and the outgroup (Hall et al. 2009). Interventions aimed at changing perceptions of the category structures of the outgroup most often focus on changing negative outgroup stereotypes and perceptions of outgroup homogeneity. For example, Johnson et al. (2013) used descriptions of counter-stereotypical outgroup exemplars, and Vezzali (2017) exposed participants to positive meta-stereotypes about the outgroup, whereas Brauer & Er-Rafiy (2011) attempted to increase the perceived variability of the outgroup by highlighting the outgroup's subgroups and diversity of opinion and characteristics, or by asking individuals to think about differences among outgroup members.

We identified 59 experimental studies within 43 papers from the past decade that tested social categorization interventions, yielding a sizeable overall average meta-analytic effect size d of 0.37 (SE = 0.05). Forty-two of these studies were conducted in the lab, which produces a larger average effect ($d = 0.44$, SE = 0.07) compared to the 16 online experiments ($d = 0.22$, SE = 0.07). We also find that social categorization interventions are most likely to measure explicit attitudes as an outcome ($d = 0.36$, SE = 0.07, $N = 45$), though implicit attitudes are also measured ($d = 0.34$, SE = 0.08, $N = 12$). If we restrict our attention to studies with treatment samples of 78 or over, we are left with only 5 studies ($d = 0.31$, SE = 0.15).

Entertainment. Entertainment is a category of interventions that seek to leverage the power of narrative or artistic transportation to overcome natural human tendencies to counterargue messages or resist persuasion attempts. Narrative transportation refers to a psychological phenomenon whereby individuals are carried away by a story, causing them to let down their tendency to question, critique, or counterargue (Green & Brock 2000). In the past decade, entertainment interventions have tested interactive narratives that allow individuals to participate in the construction of stories about outgroups (Parrott et al. 2017), films made by and for Black audiences (Eno & Ewoldsen 2010), pro-integration music lyrics (Greitemeyer & Schwab 2014), and entertainment education that incorporates educational messages about prejudice into an entertaining storyline of a soap opera or film (Murrar & Brauer 2018, Paluck & Green 2009).

We identified only 12 studies in the past decade that used entertainment interventions.[3] Of these 12, the average meta-analytic effect is quite strong ($d = 0.43$, SE = 0.07). The studies mostly measured explicit attitudes, and only four studies measured outcomes after some time passed. Five entertainment studies had a treatment sample of 78 participants or more, and these studies reported a promising average effect size d of 0.38 (SE = 0.07).

---

[3]Note that this count excludes interventions that use entertainment formats such as a film or a song for the purposes of testing other theoretical perspectives.

Face-to-face contact. Allport (1954, p. 281) first proposed that contact between members of groups may reduce prejudice given "equal status contact between majority and minority groups in the pursuit of common goals. The effect is predicted to be enhanced if this contact is sanctioned by institutional supports (i.e., by law, custom, or local atmosphere), and provided it is of a sort that leads to the perception of common interests and common humanity between members of the two groups." This hypothesis has attracted an enormous amount of social science attention (Lemmer & Wagner 2015, Paluck et al. 2019, Pettigrew & Tropp 2006). In the past decade, studies of face-to-face contact have been less numerous compared to more recent extensions of the hypothesis, described below, which rely on simulated intergroup contact. The group of experiments in our current review numbered only 28 studies across 27 manuscripts.

These studies manipulated intergroup contact across a range of different interventions. One experiment randomized ethnic minority and ethnic majority Norwegian soldiers to room with one another (or not) during basic training (Finseraas & Kotsadam 2017), another randomized criminology students to have contact with individuals incarcerated for serious crimes (Boag & Wilson 2014), and yet another randomly assigned Jewish and Arab Israelis to meet one another on peace encounters (Yablon 2012).

Across the 28 experimental studies randomizing face-to-face contact over the last decade, the average effect d is 0.28 (SE = 0.05). The studies can be divided roughly between those examining impacts on attitudes regarding race and ethnicity ($d = 0.10$, SE = 0.07, $N = 9$) and those examining prejudice toward LGBT individuals ($d = 0.22$, SE = 0.10, $N = 5$). Other prejudices were rarely targeted by experimental interpersonal contact over the last decade, and the majority of studies took place in a middle school, high school, or college setting. Seventeen manuscripts reported on studies from the lab, and ten reported from the field. The group of 8 studies that measured outcomes at least one day after the intervention reported a reduced effect size d of 0.25 (SE = 0.13). We return to some of these longer-term studies in our discussion of landmark research.

Extended and imagined contact. Extended and imagined contact are intervention approaches derived from the original contact hypothesis, intended for contexts in which face-to-face contact between members of different groups is deemed difficult or impossible (e.g., in active conflict zones). For these contexts, scholars hypothesized that similar positive attitudinal and behavioral outcomes could be achieved by indirect methods of contact, such as second-hand or imagined contact. The theory of second-hand, or extended, contact (Wright et al. 1997) is akin to theories of source credibility (Hocevar et al. 2017), insofar as it posits that learning about an ingroup member's friendship with an outgroup member will reduce prejudice toward that outgroup. Over the past decade, the majority of studies testing the extended contact hypothesis used fictional friends or characters in books or movies that belong to the same ingroup as the audience member to test whether the fictional character's contact with an outgroup member would reduce prejudice. In a study of this kind that spawned many subsequent investigations, Cameron et al. (2007) assigned children to be read stories in which able-bodied children befriended children with disabilities (treatment) or without disabilities (control).

Supposing that the benefits of contact may be achieved without actual contact, the imagined contact hypothesis further suggests that the mere exercise of imagining a positive interaction with an outgroup member should reduce prejudice. In a prototypical imagined contact intervention, participants are asked to take a few minutes to imagine themselves meeting a member of another group. Often, to enrich the imagery that is brought to mind, they are also asked to imagine the outgroup representative's appearance or a positive conversation with the person (e.g., Husnu & Paolini 2019, Turner et al. 2007, West & Bruckmüller 2013).

We found 137 studies of extended or imagined contact over the past decade from 101 manuscripts. The average meta-analytic effect size of these studies is 0.37 (SE = 0.03). The majority of these interventions focus on imagined contact with someone of another race or ethnicity ($d = 0.39$, SE = 0.08, $N = 36$), someone with a different ability ($d = 0.33$, SE = 0.05, $N = 33$), or with immigrants, asylum seekers, or refugees ($d = 0.53$, SE = 0.08, $N = 19$). Although intended for contexts, like conflict zones, where direct contact is impossible, this theory is almost never tested in those settings. The vast majority of these studies (91, in 70 manuscripts) report data from a college setting, where the average effect size is 0.36 (SE = 0.03). In younger populations the effect appears to be somewhat higher, ranging from a low of 0.49 (SE = 0.05, $N = 10$) among elementary students to a high of 0.62 (SE = 0.30, $N = 5$) among preschool to kindergarten children. The majority of all studies (123 studies in 93 manuscripts) measure explicit attitudes as an outcome ($d = 0.39$, SE = 0.04); for the most part (123 studies in 91 manuscripts), they measure the outcome immediately following an imagined positive interaction or an account of an ingroup member who is friends with an outgroup member ($d = 0.37$, SE = 0.04). The demand characteristics of interventions that instruct participants to imagine positive interactions (and the like) raise concerns about validity that are seldom addressed directly in this body of research, underscoring a need for more unobtrusive or delayed outcome measurement.

Another concern regarding the imagined and extended contact literature is the recent failure to replicate the very large effect size of one of the classic imaginary contact studies (Husnu & Crisp 2010). In contrast to the original study's reported effect size of 0.86 (SE = 0.36), the Many Labs replication project (Klein et al. 2014) found an average effect size of only 0.10 (SE = 0.05). Similarly, when we restrict our sample to the 18 studies with treatment samples of 78 or more, excluding the Many Labs replication, we find an average effect for imaginary and extended contact of 0.12 (SE = 0.04).[4]

TRENDS IN THEORY AND ASSESSMENT

We now turn our attention to broader trends in the development of the scholarly literature. This section offers observations about trends in measurement and the types of theories and hypotheses that are prioritized as of late when constructing interventions.

Outcome Measurement: Implicit Versus Explicit Prejudice Reduction

One of the many long-standing debates that suffuse the prejudice literature concerns how to assess changes in prejudice. By 2007, the earliest year considered in this review, scholars continued to debate about whether explicit survey measures accurately convey respondents' true behavioral predispositions (Jost et al. 2009, Kurdi et al. 2018, LaPiere 1934, Oswald et al. 2013), especially when respondents sense that overt expressions of outgroup hostility will be frowned upon (Roese & Jamieson 1993). Implicit bias was originally conceptualized as a related but separate type of bias, although subsequent usages of the term varied (see Brownstein et al. 2019). Among some prejudice reduction scholars, implicit prejudice is used to indicate a subtle type of prejudice that individuals do not consciously reflect upon (e.g., Sekaquaptewa et al. 2003). In its original scholarly definition, implicit prejudice is a response that bypasses conscious awareness and is captured with tools like

---

Supplemental Material >

[4]In the **Supplemental Appendix**, we include a recalculation of the imagined and extended contact average effect size including the Many Labs results as well as those of two other omnibus reviews (Lai et al. 2016, 2014); the addition of these studies, which are not indexed by our search, produces no change in our findings.

the Implicit Association Test and other timed-response measures (Greenwald & Krieger 2006, Greenwald et al. 2003). We coded which studies used this latter definition, along with response latency measures.

Recent scholarly reviews have asked whether implicit bias is related to behavior (Forscher et al. 2019, Kurdi et al. 2018) and if implicit prejudice can be reduced (Forscher et al. 2019; Lai et al. 2013, 2014). These reviews have not reported results that justify the widespread enthusiasm for implicit bias interventions in the real world; neither does our current investigation. We find that only 17% of experiments evaluating interventions in the last decade included implicit measures. On average, these interventions reported moderate effects on implicit bias ($d = 0.35$, SE $= 0.05$). Like previous reviews (Forscher et al. 2019, Kurdi et al. 2018), ours finds that many of these studies were underpowered (17 of them had 25 or fewer participants in the treatment group). We also find symptoms of publication bias insofar as studies in the smallest quintile of study sample size report an average d of 0.77 (SE $= 0.18$, $N = 11$) compared to a d of 0.26 found by studies with 78 or more participants in the treatment group (SE $= 0.10$, $N = 9$).

Importantly, we found 34 studies that assess the effects of an intervention on both implicit and explicit attitudes. Across studies, there appears to be no correlation ($r = 0.02$) between estimated effects on implicit prejudice and estimated effects on explicit prejudice. This finding is consistent with the notion that the two measures gauge distinct (and largely unrelated) response tendencies. The question is whether movement in implicit outcomes coincides with movement in behavioral outcomes; unfortunately, behavioral outcomes were only measured in exactly two experiments that also measured implicit outcomes.

Thus, a fair assessment of our data on implicit prejudice reduction is that the evidence is thin. Together with the lack of evidence for diversity training, these studies do not justify the enthusiasm with which implicit prejudice reduction trainings have been received in the world over the past decade. Perhaps due to the popularity of the term itself, implicit bias trainings are now regularly suggested in the wake of lawsuits or investigative findings that reveal unequal outcomes for non-White people, women, or LGBT individuals (Forscher & Devine 2017). The prejudice reduction literature reveals that the effects of any kind of intervention (not just implicit bias trainings) on implicit prejudice tend to be small in studies that can estimate the effect with precision and that it remains unknown whether interventions that reduce implicit bias also reduce prejudiced behavior. Given the recent literature suggesting that an emphasis on unconscious prejudice in explanations for discrimination and inequality may be interpreted as exonerating perpetrators of prejudiced behavior from responsibility (Daumeyer et al. 2019), it may be time to ramp up our investigation of attempts to reduce implicit prejudice while pausing the application of implicit prejudice interventions.

Light Touch Interventions

One noteworthy change in scholarly focus that occurred since our last review is the advent of light touch interventions. Although the precise demarcation of such interventions is a judgment call, we coded light touch interventions as those that are easy to implement, brief, and cheap. An example of a light touch intervention, described by Shih et al. (2013), is one in which participants were instructed to imagine how a fictional character felt as they watched a 3-minute film clip. By contrast, an intervention using similar materials that we did not consider to be light touch is the one discussed by Bruneau & Saxe (2012). In this study, Mexican immigrants and White Americans in Arizona were asked to reflect on and then write about their own experiences and the experiences of members of the other group, based on real interactions with one another on a simultaneous video and text chat. Interestingly, although the two literatures rarely overlap, light

touch prejudice reduction interventions sometimes resemble the "nudge" interventions from the behavioral science literature, in that they are minimal interventions that suggest what individuals might choose to attain as a positive outcome while preserving individuals' autonomy to behave as they please (Thaler & Sunstein 2009). Like nudges, these light touch interventions typically focus on individual thought and choice, but unlike nudges, they rarely tweak structural features of the environment like institutional rules or reminders. Of the 418 studies in our meta-analytic database, 76% were coded as testing one or more light touch interventions.

The emphasis on this type of intervention raises a number of questions. First, although they are nearly costless to implement, do subtle treatments have lasting effects? One way to think about this policy question is to note that even "free" interventions carry a cost of deflecting attention away from other, more time-intensive or expensive interventions. The prejudice reduction literature is largely silent on whether light touch interventions durably reduce prejudice. Only 8% of the 319 light touch intervention studies measure outcomes at least one day after treatment, and 1% measure outcomes a month later.

Thus, although a meta-analysis that focuses solely on light touch interventions suggests that they work well ($d = 0.35$, SE = 0.02), examining light touch experiments in the top quintile of sample size reveals a much smaller average effect ($d = 0.16$, SE = 0.02). Additionally, the evidence base remains thin regarding the broader theoretical claim (Walton 2014) that light touch interventions set in motion changes in perceptions, beliefs, attitudes, or behaviors that propel larger and long-term changes in prejudiced behavior and psychology (see Lai et al. 2016 for rare evidence showing that the effects of brief interventions to change implicit racial preferences do not endure for days and sometimes even hours).

Another theoretical trend we note is combining theoretical perspectives in a manner we coded as theoretical synergy (see the sidebar titled Theoretical Synergy in Creating Prejudice Reduction Interventions); the reader is referred to the **Supplemental Appendix** for further notes on audit studies and personalization of interventions.

LANDMARK STUDIES

Too often, research syntheses focus solely on estimating effect sizes, regardless of whether the treatments are realistic, the outcomes are assessed unobtrusively, and the key features of the

THEORETICAL SYNERGY IN CREATING PREJUDICE REDUCTION INTERVENTIONS

One noteworthy theoretical development is the growing interest in synergies between prejudice reduction theories. The notion of synergy refers to the idea that two (or more) theory-inspired interventions work especially well when delivered in tandem.

Recent research tests theoretical synergy in two ways. The first approach combines two theories of prejudice reduction into a single intervention but does not compare the synergistic effect to the discrete effects of each approach. The second is a more ambitious approach whereby the combination of treatments is built into the experimental design, guided by an underlying theory (or prior evidence) about how interventions interact. This allows researchers to gauge the effects of each intervention separately as well as the interactive influence of both interventions at once. The downside of this approach is that it requires a large number of participants, given that a well-powered interaction effect requires roughly 16 times as much data as a well-powered study of a single experimental factor. The average N in our database of synergy studies is just 145, which seems too few to reliably detect synergies. Thus, extra caution is required when interpreting results, pending replication.

experiment are presented in a transparent manner. Here we focus on what we term landmark studies, which are studies that are exceptionally well-designed and executed (regardless of what they discover). These studies provide a glimpse of what a meta-analysis would reveal if we could weight studies by quality as well as quantity. Not only are they interesting studies in their own right, but they also offer insights into how researchers can overcome impediments that have thus far held the literature back. We selected these landmark studies because they shed light on how the literature might step outside of its Western focus, how it might collect meaningful behavioral evidence, how surveys might track opinion unobtrusively in the wake of field interventions, and how research might exemplify transparency and data openness. These are studies that every psychologist should know and learn from.

Face-to-Face Intergroup Contact

Three landmark papers stand out for the intricate experimental designs they deploy to assess intergroup contact in the theoretically important domain of team sports and education.

Two studies randomized participants to experience cooperative intergroup contact on a sports team—one in India (Lowe 2020), in a cricket league created for the purpose of the study for men of different castes, and the other in Iraq (Mousa 2020), with an all-Christian soccer league integrated with Muslims at the initiative of the investigator. Both studies creatively measured a range of outcomes at the conclusion of the leagues' season: Lowe collected survey measures of trust and of friendship nominations, as well as measures of discrimination when players were given the opportunity to trade valuable goods with one another; Mousa also assessed attitudes along with behaviors like attending a post-Ramadan meal or voting for a Muslim non-teammate to receive a sportsmanship award.

Lowe found that interaction with different-caste individuals in the same team had generally positive effects on his prejudice-related outcomes, while the effect of adversarial contact was nuanced—mixed at best, and in a few cases arguably negative. Mousa's attitudinal measures of prejudice were unaffected by within-team contact, while some behavioral measures—training with Muslim players months later or voting for a Muslim for the sportsmanship award—showed a sizable reduction following the intergroup treatment. Unlike Lowe, Mousa found that the effects were strongest for successful teams.

In Scacco & Warren's (2018) study, there was no apparent intergroup intervention of which Nigerian participants were aware. Instead, the researchers randomly assigned Christian and Muslim applicants to computer classes in a manner that unobtrusively promoted two kinds of contact: classroom (random assignment to ethnically mixed or homogeneous classmates and instructors) and student partner (random assignment to an ingroup or outgroup partner) for the 16-week (64-hour) course. As in the soccer study, the findings suggest that contact can change discriminatory behavior even when attitudes remain unchanged. The authors found no evidence that randomly induced contact reduced prejudiced attitudes, increased knowledge about the outgroup, or decreased anxiety about interacting with outgroup members. However, participants who experienced intergroup contact showed a markedly reduced tendency to discriminate against outgroups in behavioral games.

Despite decades of research on intergroup contact, the findings described above are novel: It is still rare to find a systematic test of different kinds of contact (such as collaborative versus adversarial) or impacts (on attitudes versus behaviors) in the same study (Paluck et al. 2019). All three articles are notable for their transparency, offering preregistrations and open data as well as clear policy relevance. Their interventions build on existing organizations and programs.

Tracking the Effects of Interpersonal Conversations over Time

Broockman & Kalla (2016) and Kalla & Broockman (2020) tested the effects of peer influence and perspective taking in a series of groundbreaking field experiments on transgender and undocumented immigrant prejudice. Their interventions were conducted on a grand scale under naturalistic conditions. Vast numbers of voters were visited at their homes by volunteers for political causes. The canvassers used a planned sequence of questions and comments aimed at coaxing people into taking the perspective of transgender individuals (versus control conversations). They found that conversations with canvassers at the voters' doorsteps reduced prejudice toward these groups. The studies assessed the persistence of the effects over a period of months, using a seemingly unrelated online panel survey.

These results paint a complex theoretical picture of perspective taking and prejudice. The initial study found support for the prejudice-reducing effect of conversations that involve both perspective taking and perspective giving (Broockman & Kalla 2016). Contrary to theories of contact-driven attitude change, the conversations about transgender rights worked about equally well regardless of whether the (randomly assigned) canvassers themselves were transgender or not. Kalla & Broockman (2020) disaggregated the theoretical components of the perspective-taking and perspective-giving treatment and found that they do not appear to have additive effects. As noted below, this finding suggests that the relationship between dosage and effect size is complex. Finally, these studies are exemplars of research transparency, as the analyses are guided by open planning documents and the data are public.

Using Unobtrusive Measurement to Gauge the Effects of Workplace Diversity Training

The study by Chang et al. (2019) represents a quantum leap beyond the diversity training experiments that preceded it. Working with a large multinational corporation that required its employees to complete an hour-long online diversity training course, they used a sample that is an order of magnitude larger than other diversity training randomized controlled trials. The authors' level of access to internal data allowed them to leverage a telling array of behavioral outcome measures, such as whom employees nominated for internal awards and whom they volunteered to mentor over coffee. The combination of large *N* and behavioral outcomes allowed the authors to assess treatment effects unobtrusively over time among subgroups, which led to the discovery that the main behavioral effects of diversity training were found not among White men but rather among women and minorities. The study provides a valuable template for other training studies and leaves open the question of whether face-to-face training would differ in its effects.

Normative Interventions in Social Media Platforms

Munger's (2017) experiment tested whether social sanctioning can reduce prejudiced behavior. Social norms theory suggests that peers or elites can communicate about what kinds of behaviors are *not* typical or desirable, thus establishing clear normative boundaries. Munger conducted his test on Twitter, one of the most common sites of social sanctioning. He randomly assigned Twitter bots to reproach male users who publicly and repeatedly used the N-word slur to harass other users. The bots were randomly assigned to appear as White or Black, and with many or few followers (as a marker of status). The dependent measure of this experiment was the harasser's subsequent behavior over time: Did the harasser continue to tweet this slur over the course of

subsequent months (not just immediately following the treatment)? All of these hypotheses and analyses were preregistered, and the author posted the data and code from the experiment.

Munger's findings indicate that only sanctioning from an in-group user with a high number of followers led to an immediate average reduction in harassment, which gradually declined over the next two months (although a subset of those rebuked in this way increased their level of harassment). No effects were apparent for any other type of intervenor, including White sanctioners with few followers. The theoretical implications are that norm enforcement can be influential, but only when the messenger is a credible member of the ingroup, and even then the effect of a single enforcement gradually fades. The portability of Munger's research paradigm is demonstrated by recent work that uses similar norm enforcement in the domain of Arab-language Twitter and anti-Shia slurs (Siegel & Badaan 2020).

What Do We Learn from the Landmark Studies?

One of the sobering takeaways from this body of research is that while we find that the interventions often achieve some of their goals of reducing prejudice, their effects are often limited in size, scope, or duration. All of our landmark studies find modestly sized effects, much smaller than those reported on average in the corresponding laboratory literature using theoretically similar interventions. Furthermore, some studies find effects that diminish over time (Munger 2017) or are localized to some types of participants (Chang et al. 2019).

Several find effects only on some types of outcomes but not others. One of the most interesting discoveries that emerges from these studies is the divergence between attitudes and behaviors. Although attitudes and behaviors are correlated, the prejudice reduction interventions often seem more successful at changing discriminatory behaviors than at reducing negative stereotypes or animus (Mousa 2020, Scacco & Warren 2018). Suppose this pattern of results proves to be robust in subsequent research: What would this pattern imply theoretically? Why is there no feedback from behavior change to attitude change, as implied by consistency theories? Is the lack of feedback from behaviors to attitudes a challenge to "wise" interventions premised on the idea that small changes in experience and perception can set in motion a cumulative process that leads to profound change (Lai et al. 2016, Walton 2014)?

Many theories, we note, have yet to have their landmark study; this should be something that every promising theory aspires to achieve. Not only do large, transparent studies in the field provide an especially convincing test of a theory, they also contribute to our understanding of how theoretical insights can be translated into prejudice reductions on the ground.

CONCLUSION: THE STATE OF PREJUDICE REDUCTION

By many metrics, the study of interventions designed to reduce prejudice is thriving. Over the past decade, the publication of scholarly articles has grown at an accelerating pace. The literature features active research on a broad array of theories and applied approaches. Although the literature remains predominantly US-centered, a growing number of studies are set in other countries. Several landmark prejudice reduction studies completed in the last decade are models of careful design and measurement and of ambitious theoretical testing for all of psychological science. The average effect size found by the experimental prejudice reduction literature is 0.357, a meaningful but modest shift in prejudice. One way to look at this effect size is that following an intervention, there would be 85% overlap between people who went through the intervention and people who were in the control group.

However, the current meta-analysis gives us pause regarding how much we can learn from the vast and growing collection of experiments. When we use one of the most basic metrics of

quality to look more closely at this literature, we find a dispiriting result that repeats without exception across different theoretical domains and research settings: The size of the intervention effect drops precipitously as study size increases. Specifically, the average effect size drops 48%, to 0.187, when we focus solely on the top quintile of sample sizes. On a feeling thermometer, this average effect is five times smaller than the positive shift in feelings from cool to warm observed toward gay individuals in the United States in the past two decades. The drop in average study effects suggests that the prejudice reduction studies that come to light are disproportionately those that show propitious (statistically significant) results, while studies with more disappointing or ambiguous results remain in file drawers.

The urgency of addressing prejudice requires that the research community address this publication bias with the same seriousness of purpose that led the biomedical community to adopt CONSORT standards during the 1990s. A first step, which can be achieved without changing data collection costs, is to require experiments to be preregistered in order to be eligible for publication. A second step is to require more transparent reporting of relevant outcome measures and attrition, so as to discourage selective reporting of variables or observations. In the last decade, only 8% of all prejudice reduction experiments have been accompanied by preregistration, open data, or both. Until these reforms are embraced, psychologists must read with caution and discount reported results for many prejudice reduction interventions, especially those based on small studies.

Lest one suppose that high-quality research is prohibitively expensive or difficult to execute, our review highlights several landmark studies of the past decade that were conducted (most often by doctoral students) in high-stakes contexts. Imaginative and ambitious, these studies frequently leveraged partnerships with existing organizations in order to conduct interventions at scale in settings with contentious intergroup relations. Moreover, they did so with high-dosage interventions that played out over extended periods.

For those with a taste for conducting field research, the landmark studies offer a template of useful suggestions. We put these suggestions forward for interested scholars. Find organizations in the world that are implementing interventions that draw on or address theoretical ideas about how to reduce prejudice. Working in collaboration with an implementing partner often means that the study will be larger and able to reliably detect both main and heterogeneous effects. Moreover, studies of existing real-world interventions test the theoretical ideas that can be realistically implemented on the ground. For example, many ideas regarding perspective taking are tested in the lab as possible avenues for diversity training, but the form and substance of these exercises typically changes when they are implemented in a conference room with a group of skeptical employees. The landmark studies also suggest that aspiring field experimentalists should go to one of the many places in the world (including their own backyard) where researchers have yet to conduct a rigorous evaluation of a prejudice reduction intervention. Broadening the scope of prejudice research sheds light on whether current theories travel well outside of the contexts that initially inspired them. Alternatively, select a research site based on the kinds of behavioral outcomes that are possible to observe: Drawing inspiration from Mousa (2020) and Chang et al. (2019), look for opportunities to measure meaningful behaviors that express disdain or acceptance. Field studies expand the range of meaningful behaviors investigated by the prejudice reduction literature and help move measurement beyond the survey response.

For those who prefer to conduct experiments in lab or online settings, the literature suggests at least three important ways to make illuminating contributions. The first is to sort out theoretical anomalies suggested by existing findings. A program of research on why and when prejudiced behaviors change without concomitant shifts in prejudiced attitudes would be immensely useful. Additionally, what is the general relationship between the length or intensity of the intervention

and the size of the change in prejudice? We frequently observed studies with high-dosage interventions that had little effect on prejudiced attitudes expressed weeks or months after the intervention. How should this pattern be understood in light of the large effects that are typically reported for brief, light touch laboratory and online interventions? If these interventions genuinely produce large effects that persist over time, we urgently need to understand how and why these interventions produce outsized effects so that corresponding interventions can be developed for real-world use. If not, the sooner we clear these red herrings from our path, the less scholarly attention will be diverted from more productive lines of investigation. Until this issue is sorted out, scholars cannot credibly recommend brief, nudge-like interventions to practitioners.

Second, lab and online experimentation can play an important role in accelerating the research and development process. Thanks to the level of experimental control that these settings make possible, researchers can readily manipulate the ingredients of an anti-prejudice intervention—message, messenger, and dosage—that are often difficult and costly to manipulate in a field setting. The challenge is to do so in a way that minimizes demand effects, which again militates in favor of measuring outcomes over time, rather than in the immediate wake of the intervention. A further opportunity for researchers in the lab and online is to make efficient use of academic effort and resources through sharing of intervention materials and outcome measures.

Third, research in this domain should strive to be both theoretically broad and nimble, mixing ingredients from multiple theoretical perspectives and rediscovering older theoretical propositions that were excised from the modern versions. We urge lab researchers to revive theoretical ideas that predate the recent fascination with nudge-like interventions. We found that over one-quarter of all studies in the last decade reminded individuals to think differently about outgroups, typically through a single cognitive and emotional exercise; one-third of the interventions focused on a single imagined encounter with an outgroup. Laboratory and online studies could be used to reharness the more complex and powerful psychological forces theorized by scholars interested in social norms and group dynamics, in authority, hierarchy, and the like. Such forces arguably contribute to the acquisition and expression of prejudice; we should look there, too, for ways to harness these elemental psychological forces in the service of prejudice reduction.

Current trends in this literature may not serve the practical purposes of reducing prejudice. In our 2009 review, we likened many interventions to remedies for low-grade illnesses, noting that stronger medicine is demanded by global conditions in which prejudice is expressed through violence, discrimination, and exclusion. In the present review, we identify systematic problems with the experimental literature that occlude our view on which interventions are successful. However, aspects of this literature give hope for the development of interventions that work. Part of the reason to applaud the increasingly interdisciplinary nature of this literature is that prejudice is a vast problem; we need to pull from all corners of behavioral science to develop feasible, transportable, and sustainable ways to end prejudice in our time.

DISCLOSURE STATEMENT

ACKNOWLEDGMENTS

University and the Rita Allen Foundation to E.L.P., and from the Columbia University Graduate School of Arts and Sciences to D.P.G.

LITERATURE CITED

Alimo CJ. 2012. From dialogue to action: the impact of cross-race intergroup dialogue on the development of white college students as racial allies. *Equity Excell. Educ.* 45:36–59

Allport G. 1954. *The Nature of Prejudice*. Boston, MA: Addison-Wesley

Balas R, Sweklej J. 2013. Changing prejudice with evaluative conditioning. *Polish Psychol. Bull.* 44:379–83

Banakou D, Hanumanthu PD, Slater M. 2016. Virtual embodiment of White people in a Black virtual body leads to a sustained reduction in their implicit racial bias. *Front. Hum. Neurosci.* 10:601

Berthold A, Leicht C, Methner N, Gaum P. 2013. Seeing the world with the eyes of the outgroup—the impact of perspective taking on the prototypicality of the ingroup relative to the outgroup. *J. Exp. Soc. Psychol.* 49:1034–41

Bezrukova K, Spell CS, Perry JL, Jehn KA. 2016. A meta-analytical integration of over 40 years of research on diversity training evaluation. *Psychol. Bull.* 142:1227–74

Boag EM, Wilson D. 2014. Inside experience: Engagement empathy and prejudice towards prisoners. *J. Crim. Psychol.* 4(1):33–43

Brauer M, Er-Rafiy A. 2011. Increasing perceived variability reduces prejudice and discrimination. *J. Exp. Soc. Psychol.* 47:871–81

Broockman D, Kalla J. 2016. Durably reducing transphobia: a field experiment on door-to-door canvassing. *Science* 352:220–24

Brownstein M, Madva A, Gawronski B. 2019. What do implicit measures measure? *Wiley Interdiscip. Rev. Cogn. Sci.* 10:e1501

Bruneau EG, Saxe R. 2012. The power of being heard: the benefits of "perspective-giving" in the context of intergroup conflict. *J. Exp. Soc. Psychol.* 48:855–66

Cameron L, Rutland A, Brown R. 2007. Promoting children's positive intergroup attitudes towards stigmatized groups: extended contact and multiple classification skills training. *Int. J. Behav. Dev.* 31:454–66

Castillo JLÁ, Equizábal AJ, Cámara CP, González HG. 2014. The fight against prejudice in older adults: perspective taking effectiveness. *Rev. Latinoam. Psicol.* 46:137–47

Chaney KE. 2016. *Ends: the endurance, depth, and scope of confronting as a prejudice reduction strategy*. Ph.D. Thesis, Rutgers University, New Brunswick, NJ

Chang EH, Milkman KL, Gromet DM, Rebele RW, Massey C, et al. 2019. The mixed effects of online diversity training. *PNAS* 116:7778–83

Daumeyer NM, Onyeador IN, Brown X, Richeson JA. 2019. Consequences of attributing discrimination to implicit versus explicit bias. *J. Exp. Soc. Psychol.* 84:103812

Dessel AB. 2010. Effects of intergroup dialogue: public school teachers and sexual orientation prejudice. *Small Group Res.* 41:556–92

Devine PG, Forscher PS, Austin AJ, Cox WT. 2012. Long-term reduction in implicit race bias: a prejudice habit-breaking intervention. *J. Exp. Soc. Psychol.* 48:1267–78

Devine PG, Monteith MJ. 1999. Automaticity and control in stereotyping. In *Dual-Process Theories in Social Psychology*, ed. S Chaiken, Y Trope, pp. 339–60. New York: Guilford

Eno CA, Ewoldsen DR. 2010. The influence of explicitly and implicitly measured prejudice on interpretations of and reactions to Black film. *Media Psychol.* 13:1–30

Festinger L. 1962. *A Theory of Cognitive Dissonance*, Vol. 2. Palo Alto, CA: Stanford Univ. Press

Finseraas H, Kotsadam A. 2017. Does personal contact with ethnic minorities affect anti-immigrant sentiments? Evidence from a field experiment. *Eur. J. Political Res.* 56:703–22

Fiske S, Lin M, Neuberg SL. 1999. The continuum model: ten years later. In *Dual Process Theories in Social Psychology*, ed. S Chaiken, Y Trope, pp. 703–22. New York: Guilford

Forscher PS, Devine PG. 2017. Knowledge-based interventions are more likely to reduce legal disparities than are implicit bias interventions. In *Enhancing Justice: Reducing Bias*, ed. SE Redfield, pp. 303–16. Chicago: Am. Bar Assoc.

Forscher PS, Lai CK, Axt JR, Ebersole CR, Herman M, et al. 2019. A meta-analysis of procedures to change implicit measures. *J. Personal. Soc. Psychol.* 117:522–59

Forscher PS, Mitamura C, Dix EL, Cox WT, Devine PG. 2017. Breaking the prejudice habit: mechanisms, timecourse, and longevity. *J. Exp. Soc. Psychol.* 72:133–46

French AR, Franz TM, Phelan LL, Blaine BE. 2013. Reducing Muslim/Arab stereotypes through evaluative conditioning. *J. Soc. Psychol.* 153:6–9

Gaertner SL, Dovidio JF. 2000. *Reducing Intergroup Bias: The Common Ingroup Identity Model*. Hove, UK: Psychol. Press

Galinsky AD, Moskowitz GB. 2000. Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *J. Personal. Soc. Psychol.* 78:708–24

Goldstein NJ, Cialdini RB, Griskevicius V. 2008. A room with a viewpoint: using social norms to motivate environmental conservation in hotels. *J. Consum. Res.* 35:472–82

Gómez Á, LR Tropp, Vázquez A, Voci A, Hewstone M. 2018. Depersonalized extended contact and injunctive norms about cross-group friendship impact intergroup orientations. *J. Exp. Soc. Psychol.* 76:356–70

Green MC, Brock TC. 2000. The role of transportation in the persuasiveness of public narratives. *J. Personal. Soc. Psychol.* 79:701–21

Greenwald AG, Krieger LH. 2006. Implicit bias: scientific foundations. *Calif. Law Rev.* 94:945–67

Greenwald AG, Nosek BA, Banaji MR. 2003. Understanding and using the implicit association test: I. An improved scoring algorithm. *J. Personal. Soc. Psychol.* 85:197–216

Greitemeyer T, Schwab A. 2014. Employing music exposure to reduce prejudice and discrimination. *Aggress. Behav.* 40:542–51

Gross JJ. 2013. *Handbook of Emotion Regulation*. New York: Guilford

Gurin P, Nagda BRA, Zuniga X. 2013. *Dialogue Across Difference: Practice, Theory, and Research on Intergroup Dialogue*. New York: Russell Sage Found.

Hall NR, Crisp RJ, Suen Mw. 2009. Reducing implicit prejudice by blurring intergroup boundaries. *Basic Appl. Soc. Psychol.* 31:244–54

Halperin E, Porat R, Tamir M, Gross JJ. 2013. Can emotion regulation change political attitudes in intractable conflicts? From the laboratory to the field. *Psychol. Sci.* 24:106–11

Heitland K, Bohner G. 2010. Reducing prejudice via cognitive dissonance: Individual differences in preference for consistency moderate the effects of counter-attitudinal advocacy. *Soc. Influ.* 5:164–81

Hocevar KP, Metzger M, Flanagin AJ. 2017. Source credibility, expertise, and trust in health and risk messaging. *Oxford Research Encyclopedia of Communication*. **https://oxfordre.com/communication/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-287?print=pdf**

Husnu S, Crisp RJ. 2010. Imagined intergroup contact: a new technique for encouraging greater inter-ethnic contact in cyprus. *Peace Confl. J. Peace Psychol.* 16:97–108

Husnu S, Paolini S. 2019. Positive imagined contact is actively chosen: exploring determinants and consequences of volitional intergroup imagery in a conflict-ridden setting. *Group Process. Intergroup Relat.* 22:511–29

Johnson DR, Jasper DM, Griffin S, Huffman BL. 2013. Reading narrative fiction reduces Arab-Muslim prejudice and offers a safe haven from intergroup anxiety. *Soc. Cogn.* 31:578–98

Jost JT, Rudman LA, Blair IV, Carney DR, Dasgupta N, et al. 2009. The existence of implicit bias is beyond reasonable doubt: a refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Res. Organ. Behav.* 29:39–69

Kalev A, Dobbin F, Kelly E. 2006. Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *Am. Sociol. Rev.* 71:589–617

Kalla JL, Broockman DE. 2020. Reducing exclusionary attitudes through interpersonal conversation: evidence from three field experiments. *Am. Political Sci. Rev.* 114:410–25

Kawakami K, Phills CE, Steele JR, Dovidio JF. 2007. (Close) distance makes the heart grow fonder: improving implicit racial attitudes and interracial interactions through approach behaviors. *J. Personal. Soc. Psychol.* 92:957–71

Klein RA, Ratliff KA, Vianello M, Adams RB, Bahník Š, et al. 2014. Investigating variation in replicability: a "many labs" replication project. *Soc. Psychol.* 45:142–52

Kraimer M, Bolino M, Mead B. 2016. Themes in expatriate and repatriate research over four decades: What do we know and what do we still need to learn? *Annu. Rev. Organ. Psychol. Organ. Behav.* 3:83–109

Kurdi B, Seitchik AE, Axt J, Carroll T, Karapetyan A, et al. 2018. Relationship between the implicit association test and intergroup behavior: a meta-analysis. *Am. Psychol.* 74:569–86

Lai C, Hoffman K, Nosek B. 2013. Reducing implicit prejudice. *Soc. Personal. Psychol. Compass* 7:315–30

Lai C, Marini M, Lehr S, Cerruti C, Shin JE, et al. 2014. Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *J. Exp. Psychol. Gen.* 143:1765–85

Lai C, Skinner A, Cooley E, Murrar S, Brauer M, et al. 2016. Reducing implicit racial preferences: II. Intervention effectiveness across time. *J. Exp. Psychol. Gen.* 145:1001–16

LaPiere RT. 1934. Attitudes versus actions. *Soc. Forces* 13:230–37

Latu IM. 2010. *Reducing automatic stereotype activation: mechanisms and moderators of situational attribution training*. PhD Thesis, Ga. State Univ., Atlanta

Lehmiller JJ, Law AT, Tormala TT. 2010. The effect of self-affirmation on sexual prejudice. *J. Exp. Soc. Psychol.* 46:276–85

Lemmer G, Wagner U. 2015. Can we really reduce ethnic prejudice outside the lab? A meta-analysis of direct and indirect contact interventions. *Eur. J. Soc. Psychol.* 45:152–68

Li J, Fan Y, Zhong HQ, Duan XL, Chen W, et al. 2019. Effectiveness of an anti-stigma training on improving attitudes and decreasing discrimination towards people with mental disorders among care assistant workers in Guangzhou, China. *Int. J. Ment. Health Syst.* 13:1

Lowe M. 2020. *Types of contact: a field experiment on collaborative and adversarial caste integration*. CESifo Work. Pap. No. 8089, CESifo, Munich, Ger.

Mendoza SA, Gollwitzer PM, Amodio DM. 2010. Reducing the expression of implicit stereotypes: reflexive control through implementation intentions. *Personal. Soc. Psychol. Bull.* 36:512–23

Moher D, Liberati A, Tetzlaff J, Altman DG. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann. Intern. Med.* 151:264–69

Mousa S. 2020. Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq. *Science* 369:866–70

Munger K. 2017. Tweetment effects on the tweeted: experimentally reducing racist harassment. *Political Behav.* 39:629–49

Murrar S, Brauer M. 2018. Entertainment-education effectively reduces prejudice. *Group Process. Intergroup Relat.* 21:1053–77

Neto F, da Conceiçao Pinto M, Mullet E. 2016. Can music reduce anti-dark-skin prejudice? A test of a cross-cultural musical education programme. *Psychol. Music* 44:388–98

Oh SY, Bailenson J, Weisz E, Zaki J. 2016. Virtually old: embodied perspective taking and the reduction of ageism under threat. *Comput. Hum. Behav.* 60:398–410

Olson MA, Fazio RH. 2006. Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personal. Soc. Psychol. Bull.* 32:421–33

Oswald FL, Mitchell G, Blanton H, Jaccard J, Tetlock PE. 2013. Predicting ethnic and racial discrimination: a meta-analysis of IAT criterion studies. *J. Personal. Soc. Psychol.* 105:171–92

Paluck EL. 2006. Diversity training and intergroup contact: a call to action research. *J. Soc. Issues* 62:577–95

Paluck EL. 2009a. Reducing intergroup prejudice and conflict using the media: a field experiment in Rwanda. *J. Personal. Soc. Psychol.* 96:574–87

Paluck EL. 2009b. What's in a norm? Sources and processes of norm change. *J. Personal. Soc. Psychol.* 96:594–600

Paluck EL, Green DP. 2009. Prejudice reduction: What works? A review and assessment of research and practice. *Annu. Rev. Psychol.* 60:339–67

Paluck EL, Green SA, Green DP. 2019. The contact hypothesis re-evaluated. *Behav. Public Policy* 3:129–58

Park B, Rothbart M. 1982. Perception of out-group homogeneity and levels of social categorization: memory for the subordinate attributes of in-group and out-group members. *J. Personal. Soc. Psychol.* 42:1051–68

Parrott S, Carpentier FRD, Northup CT. 2017. A test of interactive narrative as a tool against prejudice. *Howard J. Commun.* 28:374–89

Patel SL. 2013. *Examining the influence of perceived social consensus information on weight prejudice across development*. PhD Thesis, Univ. Tex., Dallas

Perry S. 2011. *Responses to prejudice feedback on the race-implicit associations test and the role of bias awareness*. PhD Thesis, Univ. Ill., Chicago

Pettigrew TF, Tropp LR. 2006. A meta-analytic test of intergroup contact theory. *J. Personal. Soc. Psychol.* 90:751–83

Prentice D, Paluck EL. 2020. Engineering social change using social norms: lessons from the study of collective action. *Curr. Opin. Psychol. Bull.* 35:138–42

Radnitz S. 2018. Historical narratives and post-conflict reconciliation: an experiment in Azerbaijan. *Confl. Manag. Peace Sci.* 35:154–74

Robinson C. 2010. Cross-cutting messages and political tolerance: an experiment using evangelical Protestants. *Political Behav.* 32:495–515

Roese NJ, Jamieson DW. 1993. Twenty years of bogus pipeline research: a critical review and meta-analysis. *Psychol. Bull.* 114:363–75

Rosenthal R. 1979. The "file drawer problem" and tolerance for null results. *Psychol. Bull.* 86:638–41

Scacco A, Warren SS. 2018. Can social contact reduce prejudice and discrimination? Evidence from a field experiment in Nigeria. *Am. Political Sci. Rev.* 112:654–77

Sechrist GB, Milford-Szafran LR. 2011. "I depend on you, you depend on me. Shouldn't we agree?": the influence of interdependent relationships on individuals' racial attitudes. *Basic Appl. Soc. Psychol.* 33:145–56

Sekaquaptewa D, Espinoza P, Thompson M, Vargas P, von Hippel W. 2003. Stereotypic explanatory bias: implicit stereotyping as a predictor of discrimination. *J. Exp. Soc. Psychol.* 39:75–82

Sherman DK, Cohen GL. 2006. The psychology of self-defense: self-affirmation theory. *Adv. Exp. Soc. Psychol.* 38:183–242

Shih MJ, Stotzer R, Gutiérrez AS. 2013. Perspective-taking and empathy: generalizing the reduction of group bias towards Asian Americans to general outgroups. *Asian Am. J. Psychol.* 4:79–83

Siegel AA, Badaan V. 2020. #No2sectarianism: experimental approaches to reducing sectarian hate speech online. *Am. Political Sci. Rev.* 114:837–55

Simonsohn U, Nelson LD, Simmons JP. 2014. P-curve: a key to the file-drawer. *J. Exp. Psychol. Gen.* 143:534–47

Smeekes A, Verkuyten M, Poppe E. 2012. How a tolerant past affects the present: historical tolerance and the acceptance of Muslim expressive rights. *Personal. Soc. Psychol. Bull.* 38:1410–22

Smith BD, Silk K. 2011. Cultural competence clinic: an online, interactive, simulation for working effectively with Arab American Muslim patients. *Acad. Psychiatry* 35:312–16

Stell AJ, Farsides T. 2016. Brief loving-kindness meditation reduces racial bias, mediated by positive other-regarding emotions. *Motiv. Emot.* 40:140–47

Stewart BD, Payne BK. 2008. Bringing automatic stereotyping under control: implementation intentions as efficient means of thought control. *Personal. Soc. Psychol. Bull.* 34:1332–45

Tajfel H, Turner J. 1979. An integrative theory of intergroup conflict. In *The Social Psychology of Intergroup Relations*, ed. WG Austin, S Worchel, pp. 33–47. Monterey, CA: Brooks/Cole

Thaler RH, Sunstein CR. 2009. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New York: Penguin

Turner RN, Crisp RJ, Lambert E. 2007. Imagining intergroup contact can improve intergroup attitudes. *Group Process. Intergroup Relat.* 10:427–41

Vezzali L. 2017. Valence matters: positive meta-stereotypes and interethnic interactions. *J. Soc. Psychol.* 157:247–61

Villicana AJ, Rivera LM, Garcia DM. 2018. When one's group is beneficial: the effect of group-affirmation and subjective group identification on prejudice. *Group Process. Intergroup Relat.* 21:962–76

Walsh SP. 2013. *Reducing automatic stereotype activation: European and African American photos in situational attribution training*. PhD Thesis, Univ. Miss., Oxford

Walton GM. 2014. The new science of wise psychological interventions. *Curr. Dir. Psychol. Sci.* 23:73–82

West K, Bruckmüller S. 2013. Nice and easy does it: how perceptual fluency moderates the effectiveness of imagined contact. *J. Exp. Soc. Psychol.* 49:254–62

Witkowska M, Beneda M, Čehajić-Clancy S, Bilewicz M. 2019. Fostering contact after historical atrocities: the potential of moral exemplars. *Political Psychol.* 40:565–82

Woodcock A, Monteith MJ. 2013. Forging links with the self to combat implicit bias. *Group Process. Intergroup Relat.* 16:445–61

Wright SC, Aron A, McLaughlin-Volpe T, Ropp SA. 1997. The extended contact effect: knowledge of cross-group friendships and prejudice. *J. Personal. Soc. Psychol.* 73:73–90

Yablon YB. 2012. Are we preaching to the converted? The role of motivation in understanding the contribution of intergroup encounters. *J. Peace Educ.* 9:249–63

Annual Review of
Psychology

Volume 72, 2021

Contents

Indexes