Supplemental Material: Annu. Rev. Psychol. 2021. 72.
https://doi.org/10.1146/annurev-psych-071620-030619

Appendix for: Prejudice Reduction: Progress and Challenges

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

[1]Princeton University
[2]Hebrew University
[3]Columbia University

Contents

1 Meta-analytic Search and Analysis

This section describes in detail the steps by which we assembled the set of relevant studies on the topic of prejudice reduction.

1.1 Search terms and databases

Following the biomedical meta-analytic standards for our search (Moher, Liberati, Tetzlaff, & Altman, 2009), we searched for published and unpublished articles starting in 2007, the final year covered by the previous Paluck & Green (2009) review. We conducted our search in the following databases: *Scopus*, *PsycINFO*, *Web of Science*, and *Social Sciences Full Text*. To supplement this database search, we conducted a text-based search in the proprietary database *Articles+*. Furthermore, we searched the bibliographies of recent sub-area prejudice reduction meta-analyses (e.g., Bezrukova, Spell, Perry, & Jehn, 2016) for additional studies. To evaluate our final set of studies, we searched internally for well-known papers in the prejudice-reduction field from the past decade.

To identify relevant studies, we searched through the aforementioned databases for studies that attempt to either decrease prejudice or increase positive intergroup outcomes. To search for studies that aim to decrease prejudice, we paired the primary search words of "prejudice," "discrimination," "implicit bias," "explicit bias," "in-group bias," "stereotyp*," "racis*," "homophobi*," "abilis*," "islamophobi*," "transphobi*," "ageism," "outgroup hate," "dehumaniz*," "infrahumaniz*," "ingroup bias," "outgroup bias," "intergroup bias," or "intergroup anxiety", within seven words of "reduc*," or "decrease" with the qualifying terms of "quantitative," "study," "experiment*," "controlled study," "intervention," or "control group". To search for studies that aim to increase positive outcomes, we paired the primary search words of "humaniz*," "cultural sensitivity," "cultural competence," "reconcil*," "intergroup attitudes," or ("*group" within 2 words of "tolerance") all within 7 words of "improv*," "enhanc*," or "increas*" paired with the same qualifying terms as above.

The specific search conditions differed slightly from one database to another in reflection of database structure. Below we provide the exact search terms used for each database. We retrieved searches using these terms for the years 2007-2018 on November 11 and 12, 2019, and for the year 2019 on January 17 and 18, 2020.

Scopus Negative

(TITLE-ABS-KEY ((prejudice OR discrimination OR "Implicit bias" OR "Explicit bias" OR "In-group bias" OR stereotyp* OR racis* OR homophobi* OR abilis* OR islamophobi* OR transphobi* OR ageism OR "outgroup hate" OR dehumaniz* OR infrahumaniz* OR "ingroup bias" OR "outgroup bias" OR

"intergroup bias" OR "intergroup anxiety") W/7 (reduc* OR decrease)) AND TITLE-ABS-KEY ((quantitative OR study OR experiment* OR "Controlled study" OR intervention OR "Control Group"))) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (SUBJAREA , "PSYC") OR LIMIT-TO (SUBJAREA , "MEDI") OR LIMIT-TO (SUBJAREA , "SOCI") OR LIMIT-TO (SUBJAREA , "NEUR") OR LIMIT-TO (SUBJAREA , "BUSI") OR LIMIT-TO (SUBJAREA , "NURS") OR LIMIT-TO (SUBJAREA , "ECON") OR LIMIT-TO (SUBJAREA , "HEAL") OR LIMIT-TO (SUBJAREA , "MULT")) AND (LIMIT-TO (LANGUAGE , "English"))

Scopus Positive

(TITLE-ABS-KEY (((*group W/2 tolerance) OR humaniz* OR "cultural sensitivity" OR "cultural competence" OR reconcil* OR "intergroup attitudes") W/7 (improv* OR enhanc* OR increas*)) AND TITLE-ABS-KEY ((quantitative OR study OR experiment* OR "Controlled study" OR intervention OR "Control Group"))) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (SUBJAREA , "MEDI") OR LIMIT-TO (SUBJAREA , "SOCI") OR LIMIT-TO (SUBJAREA , "NURS") OR LIMIT-TO (SUBJAREA , "PSYC") OR LIMIT-TO (SUBJAREA , "BUSI") OR LIMIT-TO (SUBJAREA , "HEAL") OR LIMIT-TO (SUBJAREA , "NEUR") OR LIMIT-TO (SUBJAREA , "ECON") OR LIMIT-TO (SUBJAREA , "MULT") OR LIMIT-TO (SUBJAREA , "Undefined"))

Web of Science Negative

TOPIC: (((Prejudice OR Discrimination OR "Implicit bias" OR "Explicit bias" OR "In-group bias" OR Stereotyp* OR Racis* OR Homophobi* OR Abilis* OR Islamophobi* OR Transphobi* OR ageism OR "outgroup hate" OR dehumaniz* OR infrahumaniz* OR "ingroup bias" OR "outgroup bias" OR "intergroup bias" OR "intergroup anxiety") NEAR/7 (Reduc* OR Decrease))) AND TOPIC: (((Quantitative OR Study OR Experiment* OR "Controlled study" OR Intervention OR "Control Group"))) Refined by: DOCUMENT TYPES: (ARTICLE) AND [excluding] WEB OF SCIENCE CATEGORIES: (NURSING OR SURGERY OR VETERINARY SCIENCES OR TELECOMMUNICATIONS OR ZOOLOGY OR AUDIOLOGY SPEECH LANGUAGE PATHOLOGY OR AGRONOMY OR BIOCHEMISTRY MOLECULAR BIOLOGY OR HOSPITALITY LEISURE SPORT TOURISM OR AUTOMATION CONTROL SYSTEMS OR BIOCHEMICAL RESEARCH METHODS OR NUTRITION DIETETICS OR CHEMISTRY MEDICINAL OR COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE OR PHYSIOLOGY OR COMPUTER SCIENCE INTERDISCIPLINARY APPLICATIONS OR DEMOGRAPHY OR RADIOLOGY NUCLEAR MEDICINE MEDICAL IMAGING OR ECOLOGY OR ENDOCRINOLOGY METABOLISM OR ENGINEERING ELECTRICAL ELECTRONIC OR GEOGRAPHY OR GEOSCIENCES

MULTIDISCIPLINARY OR INFECTIOUS DISEASES OR INSTRUMENTS INSTRUMENTATION OR INTEGRATIVE COMPLEMENTARY MEDICINE OR ONCOLOGY OR OPHTHALMOLOGY OR REGIONAL URBAN PLANNING OR OTORHINOLARYNGOLOGY OR PLANT SCIENCES OR RESPIRATORY SYSTEM OR TOXICOLOGY OR WATER RESOURCES OR ACOUSTICS OR AGRICULTURE DAIRY ANIMAL SCIENCE OR AGRICULTURAL ENGINEERING OR PHARMACOLOGY PHARMACY OR COMPUTER SCIENCE INFORMATION SYSTEMS OR ANESTHESIOLOGY OR INDUSTRIAL RELATIONS LABOR OR BIODIVERSITY CONSERVATION OR MEDICINE RESEARCH EXPERIMENTAL OR BIOTECHNOLOGY APPLIED MICROBIOLOGY OR REMOTE SENSING OR BUSINESS FINANCE OR SPORT SCIENCES OR CARDIAC CARDIOVASCULAR SYSTEMS OR MEDICINE GENERAL INTERNAL). Indexes: SCI-EXPANDED, SSCI, ESCI.

Web of Science Positive

TOPIC: ((((*group NEAR/2 tolerance) OR humaniz* OR "cultural sensitivity" OR "cultural competence" OR reconcil* OR "intergroup attitudes") NEAR/7 (improv* OR enhanc* OR increas*))) AND TOPIC: (((Quantitative OR Study OR Experiment* OR "Controlled study" OR Intervention OR "Control Group"))) Refined by: DOCUMENT TYPES: (ARTICLE) AND WEB OF SCIENCE CATEGORIES: (NURSING OR EDUCATION EDUCATIONAL RESEARCH OR HEALTH CARE SCIENCES SERVICES OR ECONOMICS OR EDUCATION SCIENTIFIC DISCIPLINES OR MULTIDISCIPLINARY SCIENCES OR PSYCHOLOGY SOCIAL OR HISTORY PHILOSOPHY OF SCIENCE OR SOCIAL WORK OR INFORMATION SCIENCE LIBRARY SCIENCE OR HUMANITIES MULTIDISCIPLINARY OR INDUSTRIAL RELATIONS LABOR OR ETHICS OR LANGUAGE LINGUISTICS OR PSYCHOLOGY OR PSYCHOLOGY APPLIED OR BUSINESS OR FAMILY STUDIES OR NEUROSCIENCES OR GERIATRICS GERONTOLOGY OR ANTHROPOLOGY OR MANAGEMENT OR BEHAVIORAL SCIENCES OR POLITICAL SCIENCE OR PHILOSOPHY OR PSYCHIATRY OR PSYCHOLOGY CLINICAL OR PSYCHOLOGY MULTIDISCIPLINARY OR PSYCHOLOGY DEVELOPMENTAL). Indexes: SCI-EXPANDED, SSCI, ESCI.

PsycINFO Negative

((Prejudice OR Discrimination OR "Implicit bias" OR "Explicit bias" OR "In-group bias" OR Stereotyp* OR Racis* OR Homophobi* OR Abilis* OR Islamophobi* OR Transphobi*) W7 (Reduc* OR Decrease)) AND ((Quantitative OR Study OR Experiment* OR "Controlled study" OR Intervention OR "Control Group")) Limiters - Publication Type: All Journals, Dissertation Abstract; Language: English Search modes - Find all my search terms

PsycINFO Positive

(((*group W/2 tolerance) OR humaniz* OR "cultural sensitivity" OR "cultural competence" OR reconcil* OR "intergroup attitudes") W7 (improv* OR enhanc* OR increas*)) AND ((Quantitative OR Study OR Experiment* OR "Controlled study" OR Intervention OR "Control Group")) Limiters - English; Publication Type: All Journals, Dissertation Abstract Search modes - Find all my search terms

Social Sciences Full Text Negative

((Prejudice OR Discrimination OR "Implicit bias" OR "Explicit bias" OR "In-group bias" OR Stereotyp* OR Racis* OR Homophobi* OR Abilis* OR Islamophobi* OR Transphobi*) W7 (Reduc* OR Decrease)) AND (Quantitative OR Study OR Experiment* OR "Controlled study" OR Intervention OR "Control Group") Limiters - Publication Type: Academic Journal Narrow by Language: - english Search modes - Find all my search terms

Social Sciences Full Text Positive

(((*group W/2 tolerance) OR humaniz* OR "cultural sensitivity" OR "cultural competence" OR reconcil* OR "intergroup attitudes") W7 (improv* OR enhanc* OR increas*)) AND ((Quantitative OR Study OR Experiment* OR "Controlled study" OR Intervention OR "Control Group")) Limiters - Publication Type: Academic Journal Narrow by Language: - english Search modes - Find all my search terms

Selection of studies. Our exhaustive search yielded over 16,000 studies (see PRISMA flowchart below). The process of deciding whether a study meets eligibility criteria for inclusion was conducted by the authors based on both theoretical and methodological grounds.

1.2 Theoretical inclusion criteria

In order to identify the set of prejudice-reduction studies, we must first stipulate what we mean by prejudice. Recognizing that there is no single answer to this question, we sought a definition that would encompass standard usage by researchers working in this domain: prejudice is a negative bias or animus toward social groups and their putative members. From this definition follows the exclusion of a number of social science literatures:

1. We exclude the vast literature interested in mitigating the effects of prejudice on its targets (e.g., stereotype threat, social-belonging, self-affirmation).

2. We exclude priming and related interventions that did not explicitly aim to reduce prejudice but rather to demonstrate that prejudice is momentarily malleable.

3. We excluded studies examining interventions to reduce prejudice towards advantaged groups by disadvantaged group members.

4. We exclude work on types of intergroup biases that are typically explained by reference to a distinct set of theories. For example, we do not include prejudice toward political groups, ISIS, skinheads, or rival universities.

5. In a similar vein, and following the previous review, this review does not include gender-based prejudice, with the exception of prejudice toward transgender or genderqueer identity, due to its unique operation and distinct set of theories.

An additional criterion for study inclusion was the presence of at least one outcome assessing prejudice. Thus, we did not include studies whose outcomes focused solely on policy preferences or support, because these preferences can be shaped by more factors than just prejudice. The exclusion of policy outcomes inadvertently excluded a large percentage of the conflict reduction literature.

These restrictions nevertheless leave us with a vast number of relevant studies that investigate the expression of prejudice. Because our focus is on prejudice reduction, we restrict our attention to those studies that in some way seek to change beliefs and attitudes that contribute to prejudice as a psychological predisposition, or its expression in behavior or behavioral intentions.

1.3 Methodological grounds for exclusion

Meta-analysis presupposes that the reported standard errors properly summarize the statistical uncertainty associated with a given study. Because non-statistical 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 studies with non-experimental studies. We therefore exclude non-experimental studies from our meta-analysis. To be included, a study had to clearly indicate that random assignment was used to allocate participants (or groups) to treatment or control conditions.

One challenge in assembling a meta-analytic review is deciding whether and where to set a threshold for adequately designed and implemented research. While poring over the literature, we noticed four recurrent problems. We did not always exclude based on these characteristics unless the severity of the problem jeopardized the validity or reliability of the results. The first problem is attrition: subjects are lost to follow-up, fail to complete end-line surveys, or are excluded

for failing to pay attention or follow instructions. Attrition jeopardizes the unbiasedness of a randomized experiment, particularly if missingness operates differently among treatment and control groups. We encountered studies in which more than half of the participants who were randomly assigned to experimental conditions were missing at end-line, and we also observed studies where the rates or predictors of missingness appeared significantly different across experimental conditions.

Another frequent conundrum is whether and how to code cluster-randomized experiments, in which groups of people, rather than individuals, are randomly assigned to different experimental conditions. Cluster-randomized experiments create two complications. First, the sampling variance of many experiments often increases sharply when clusters, rather than individuals, are assigned to experimental conditions. Estimating the clustering penalty to the standard errors requires at least ten clusters (preferably many more; (Gerber & Green, 2012)). Second, conventional estimators such as regression yield biased results when small numbers of clusters of unequal size are randomly assigned. For these reasons, we excluded experiments with fewer than ten clusters.

Finally, we note that many of the studies we wished to include in the review failed to describe their analysis procedures in sufficient detail to assure us that the results were trustworthy. More generally, we encountered a disturbing lack of transparency insofar as few studies made their data and code publicly available. Many candidate studies neglected to report means, standard deviations, or standard errors in the tables or text of the paper itself. A lack of public data made it impossible for us to calculate effect sizes ourselves in order to verify our reading of the text and tables, but also in many cases, tables and text did not report treatment effects for the entire subject pool, instead selectively reporting results for certain subgroups. (One of the attributes that sets "landmark" studies apart from others is the exemplary manner in which their methods are described and replication materials are archived.) We were forced to exclude studies that did not report sufficient information in the paper itself; we strove whenever possible to include papers that reported just enough information for us to reconstruct the apparent average treatment effect and its standard error.[1]

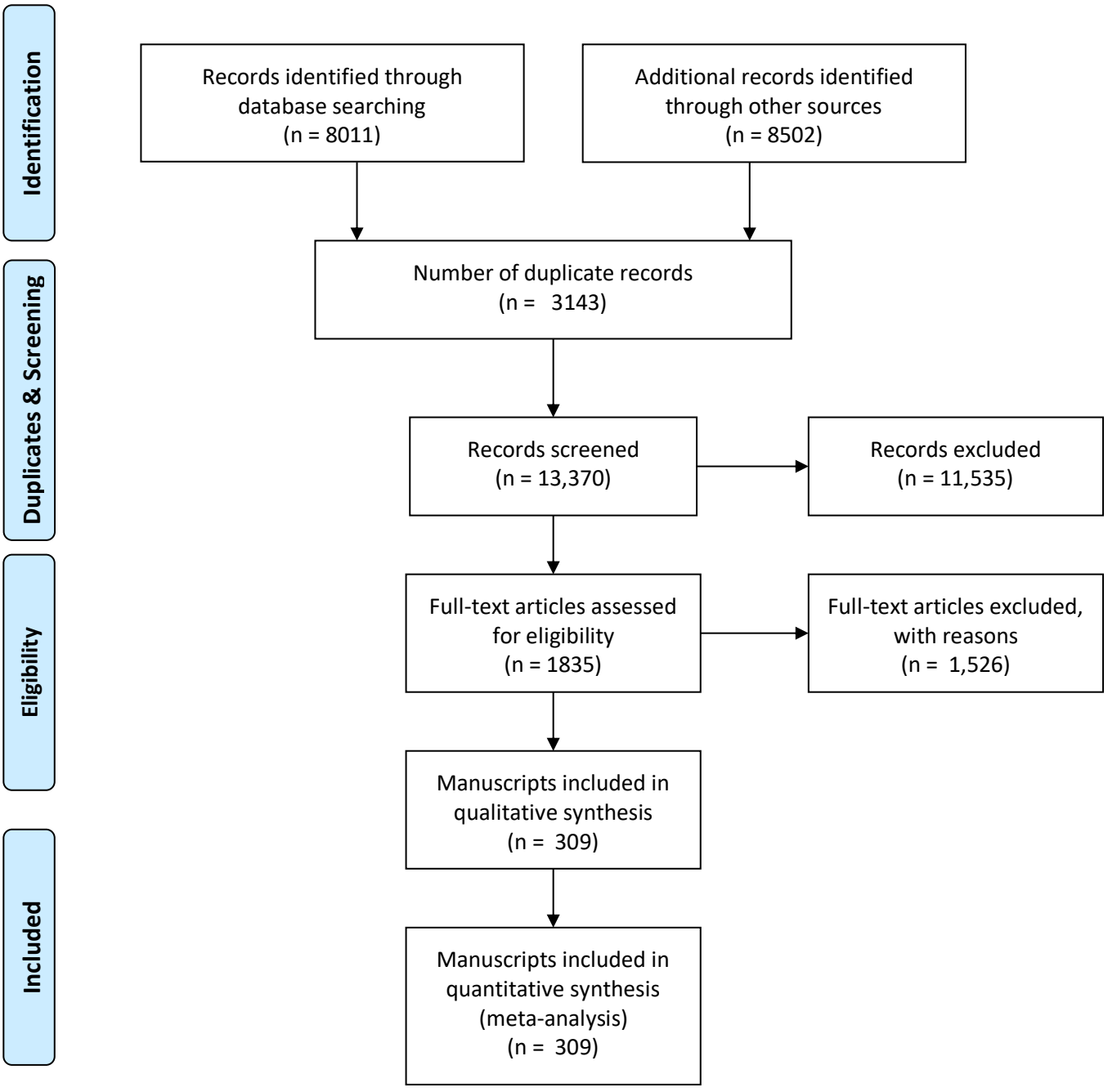
[1] Our technical appendix lists the conversion rules used to transform the published statistics into usable estimates and standard errors.

1.4 PRISMA sample documentation

PRISMA Flow Chart

Identification	4126	Main search results (through 2018)	
	742	Main search results (2019)	
	2755	Duplicates within main search	
	388	Duplicates between main and secondary search	
	8011	**Total # Main Search Results**	**"Records identified through database searching"**
	7891	Secondary database search results	
	611	Results from special searches	
	8502	**Total Secondary Search Results**	**"Additional records identified through other sources"**
	16513	Total Number of Search Results	
Screening	**3143**	**Total number of duplicates**	**"Number of duplicate records"**
	13370	**Total screened (excluding duplicates)**	**"Records screened"**
	3950	Main search results (through 2018) screened out	
	672	Main search results (2019) screened out	
	4622	Total main search results screened out	
	6913	Total secondary search results screened out	
	11535	**Total results screened out**	**"Records excluded"**
Eligibility	176	Main search results (through 2018) assessed full-text	
	70	Main search results (2019) assessed full-text	
	246	Total main search results assessed full-text	
	1589	Total secondary search results assessed full-text	
	1835	**Total results assessed full-text**	**"Full-text articles assessed for eligibility"**
	96	Main search results excluded after full-text assessment	
	1430	Secondary search results excluded after full-text	
	1526	**Total results excluded after full-text assessment**	**"Full-text articles excluded, with reasons"**
Included	**309**	**Total number of manuscripts included in final dataset (note: 416 studies)**	**"Manuscripts included in qualitative synthesis"**
	309	**Total number of manuscripts included in final dataset (note: 416 studies)**	**"Manuscripts included in quantitative synthesis (meta-analysis)"**

PRISMA 2009 Flow Diagram



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

1.5 Coding and reliability

In total we recovered 309 manuscripts from our search. Recovered papers were initially read by one of the authors who coded the high-level attributes of the paper that help us to understand how each paper contributes broadly to the prejudice reduction field. These attributes include:

1. Type of intervention approach associated with a particular theoretical test (e.g., a contact intervention vs. a peer influence intervention).

2. Classification of whether each reported study could be considered a laboratory, online, or field experiment. To asses whether a study qualifies as a field experiments, we considered four aspects of the study: (a) participants, (b) the intervention and its target, (c) the obtrusiveness of intervention delivery, and (d) the assessed response to the intervention. Laboratory and online studies were identified according the paper's description.

3. Identification of the relevant experimental conditions for which statistics needed to be coded. To do so, we identified which experimental contrasts were needed to test the study's main prejudice reduction hypotheses (excluding hypotheses about the effects of one intervention vs. another–all contrasts needed to have a pure or placebo control group so that we could determine the direction of the effect). When possible, we captured other statistical contrasts from the study design that were informative of other hypotheses in the prejudice reduction literature. For example, in Vezzali, Stathi, Crisp, and Capozza (2015), the authors set out to test the interactive effect of direct and imagined contact–a contrast that we captured for the meta-analysis (Vezzali et al., 2015). Their design also allowed us to capture the main treatment effects of direct and imagined contact separately. Because each could be contrasted to a control condition, we thus harvested the interactive and the two main effects from this paper, following the study's hypothesis and field-wide hypotheses about the positive effects of the two types of contact

4. Selection of outcomes of interest to code for each study. Because many studies reported several outcome measures. We captured up to five outcomes for each study. We always recorded the results of a behavioral and an implicit outcome, if reported. In addition, we recorded the results for up to three additional outcomes, including explicit attitudes, emotions, norms, and behavioral intentions. In cases where studies reported more relevant outcomes than we could take, we took the outcomes assigned most priority by the authors of the study (i.e., those featured in the abstract or featured more prominently in the manuscript).

After this phase, information from each study was recorded and then checked by at least two members of a team of 9 masters- or postdoctoral-level paid research assistants. This team coded both qualitative and quantitative attributes of the study. Quantitative coding included all statistics needed to calculate the

effect size of outcomes, information about the actual intervention (e.g., time duration), the target population (e.g., age and population context like school or work), the type of prejudice targeted (e.g. race, nationality, immigrants), the timing of the outcome measurement (e.g., immediately after the intervention delayed), and whether the study followed open science practices (e.g., preregistration, open data).

1.6 Qualitative coding

Our coding of studies' qualitative attributes included the following: whether the intervention might be considered "light touch" vs. time and resource-intensive, whether the intervention was in some ways personalized to individual attributes of participants, and whether the theoretical grounding of the intervention fused the insights of more than one theory together to create a synergy or interaction. A description of all of these codes is included in the coding manual, contained on our Dataverse site for this study.

1.7 Coding disagreements

Coding disagreements were resolved by documented discussions among the authors, which helped create a set of coding rules for difficult cases, contained on our Dataverse site for this study. Finally, all papers were double checked by at least one person (either a research assistant or an author) to verify both qualitative and quantitative information recorded.

1.8 Final sample

The coding process led to a winnowing of usable studies that were either off topic, non-experimental, methodologically suspect, or described in insufficient detail to allow for coding.

From our original set of approximately 16,000 manuscripts, we settled on a final set of 309 (representing 416 studies).

2 Meta-Analytic Procedures

2.1 Robustness checks for overall estimate

. The literature on prejudice reduction presents a number of technical challenges for meta-analysis. First, a given study often presents more than one outcome measure. Ignoring this complication and naively declaring each estimate to be a distinct empirical result both accords more weight to studies with more

outcome measures and exaggerates the N of subjects in such studies (since the same subjects are counted repeatedly each time they contribute to a measured outcome). Second, even if one were to distill each study down to a single outcome measure or summary, there remains the problem of analyzing multiple studies within the same article (or dissertation), which are unlikely to be independent draws from an underlying data generating process. Methodological remedies vary, and we have settled on an approach that seems to be quite robust to alternative specifications (see Appendix Table 1). To distill studies down to a single estimate, we average their reported Cohen's d for each outcome and average their estimated sampling variance. These distilled estimates (one per study) are then used as inputs to a random effects meta-analysis. To address the issue of non-independence across studies from the same article, we cluster this random effects estimate's standard error by article. The data and replication code may be found at the Dataverse site for this study.

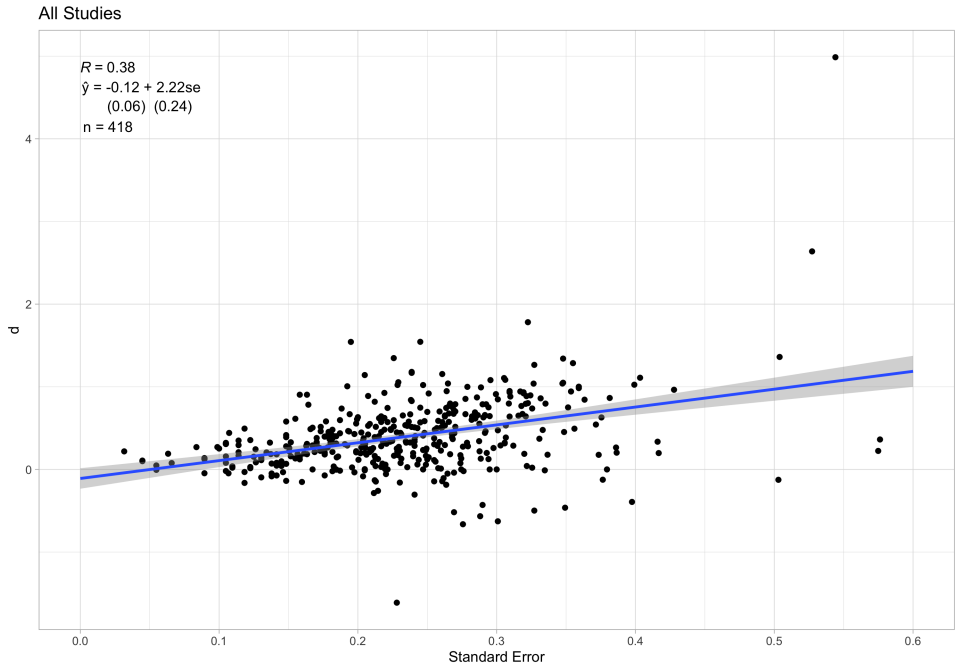
The random effects meta-analysis results reported in Appendix Table 1 show the average treatment effect across all prejudice-reduction experiments from 2007-2019. The table also provides a sense of the robustness of the results across different statistical procedures. The first row of the table (approach 1) reports the results of a naive analysis of all estimated treatment effects, ignoring the fact that some of these estimates come from the same study and are therefore not independent. The next row (approach 2)shows the same estimate but correcting its standard error by clustering at the level of the study. Even this approach is somewhat misleading because it ignores the fact that the N from a given study is double counted if that study produces two effect estimates. The third row of the table (approach 3) reports our preferred estimation approach, which eliminates double-counting by averaging estimates from each study before meta-analyzing the study-level estimates. For purposes of comparison, we show this estimate with a naive standard error that ignores the fact that the estimates from different studies may be clustered by the manuscript in which the studies are reported. Finally, the fourth row of the table (approach 4) represents our reported estimate that clusters standard errors at the study level, and averages the reported d's into one estimate per study. Interestingly, the estimates from these different approaches are quite similar, hovering around an average d of 0.34. The clustered standard errors are, as expected, much larger than the naive standard errors, but the level of sampling variability remains quite small in relation to the estimated effect size. Even the smallest z-score is larger than 15.4, allowing us to reject a null of no treatment effect at p ¡ .0001.

Appendix Table 1: **Random effects meta-analyses, by approach**

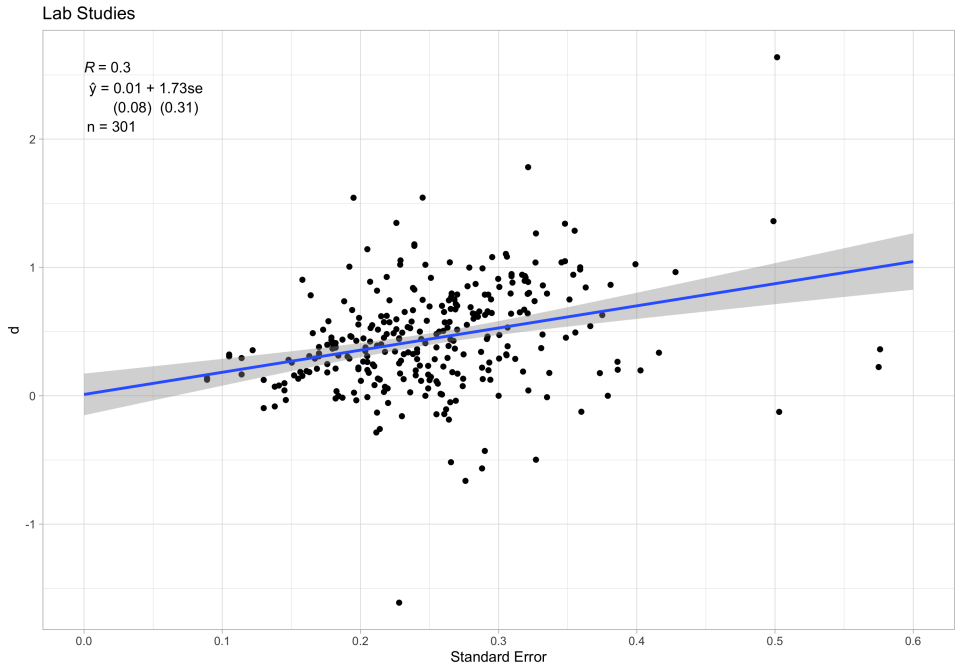
Approach	SE clustered by paper	d collapsed by study	d-estimate
1	no	no	0.326 (0.01)
2	yes	no	0.326 (0.02)
3	no	yes	0.357 (0.02)
4	yes	yes	0.357 (0.02)

2.2 Publication bias concerns

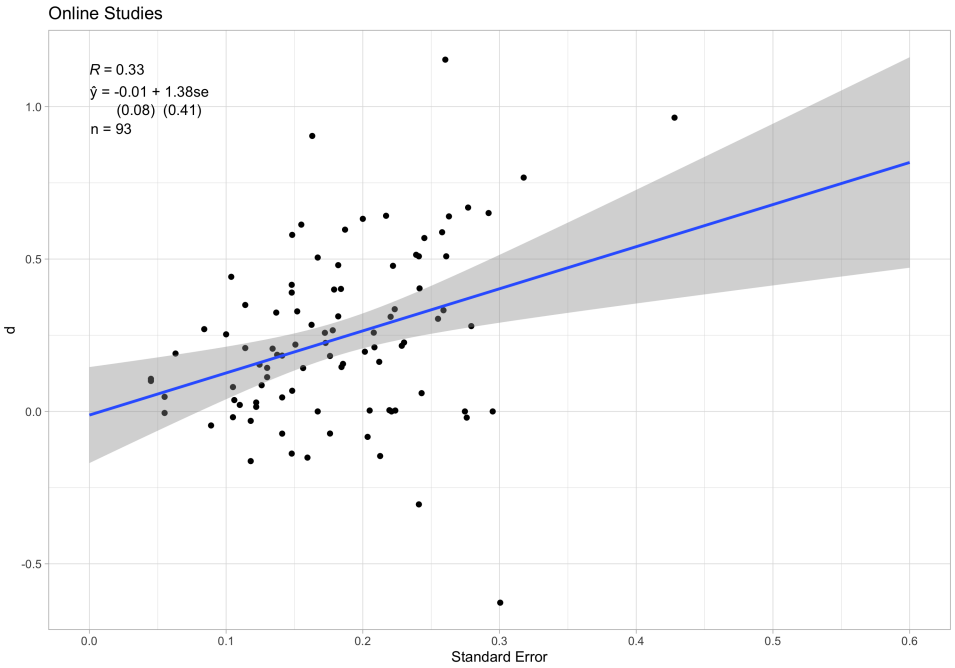
One concern articulated in the manuscript is publication bias, or the tendency for journals to favor manuscripts that report statistically significant results. The telltale sign of bias is a strong positive relationship between effect sizes and their standard errors. Put differently, in the absence of publication bias, we should obtain similar average effect estimates from small and from large studies. We do in fact see a powerful relationship of this kind in our sample: we find a strong positive relationship between standard errors and effect size. This means that studies with larger standard errors (i.e., less precise) generate large effects, whereas studies that generate precise results (i.e., smaller standard errors) tend to produce much weaker effects (i.e., smaller d's). We find this relationship in all types of studies (i.e., lab, online, and field), suggesting that publication bias is not restricted to one methodology.



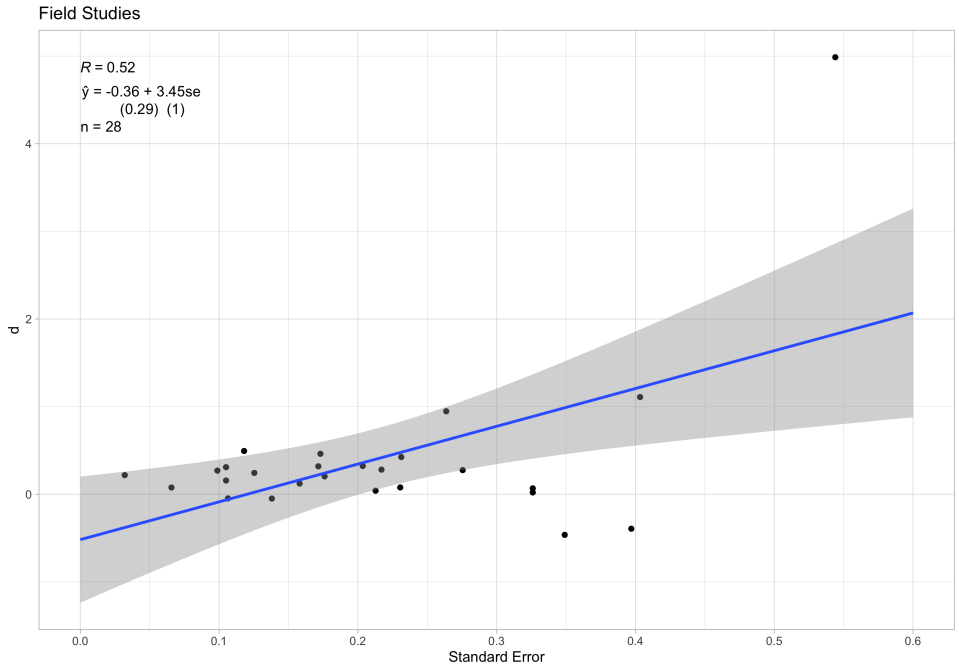
Appendix Figure 1: **Relationship between standard error and effect size for all studies**



Appendix Figure 2: **Relationship between standard error and effect size for lab studies**

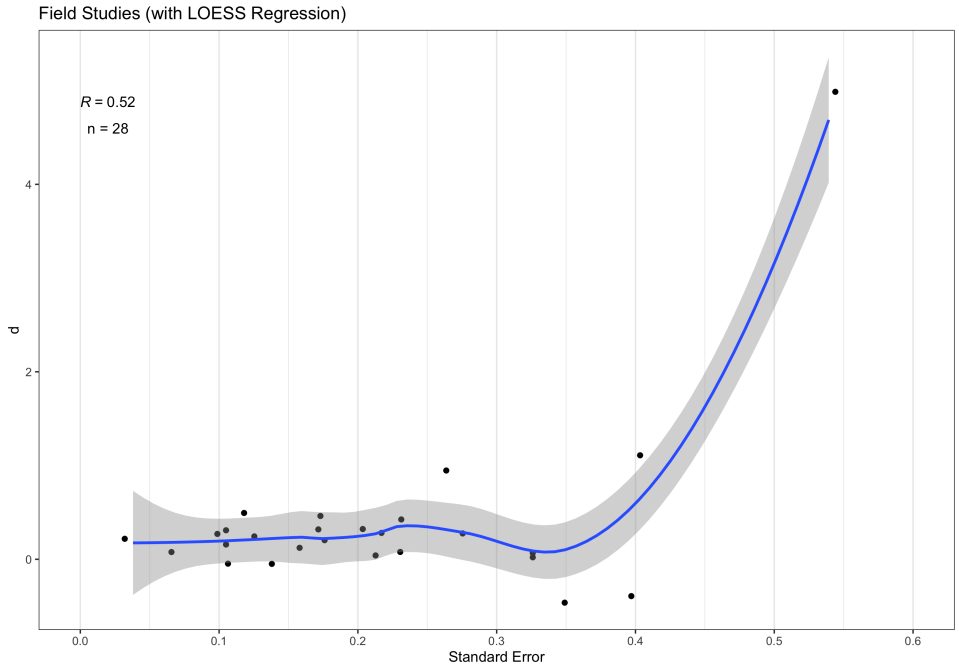


Appendix Figure 3: **Relationship between standard error and effect size for online studies**



Appendix Figure 4: **Relationship between standard error and effect size for field studies**

Given the patently non-linear nature of this relationship, we also present the data with a plotted loess regression, which shows an approximately linear relationship for most data points and one large outlier (Chongruksa, Prinyapol, Wadeng, & Padungpong, 2010).



Appendix Figure 5: **Relationship between standard error and effect size for field studies (LOESS regression)**

2.3 Robustness Check: Adding Klein et al., (2014), and Lai et al., (2014; 2016).

We conduct a robustness check by including three large-scale replication papers - Klein et al., (2014) and Lai et al., (2014; 2016) to our results to note whether there are any changes to our conclusions. These papers were not picked up by our original search criteria and were therefore not included in the final manuscript. When including these studies we find that they hardly change our estimates, and do not change our overall conclusions.

We first turn to Klein et al. (2014) Many Labs replication project, which replicated 13 well-known effects, including the imagined contact intervention originally reported by Husnu and Crisp (2010a). The replication included 36 samples with over 6,000 participants. When we include this effect for estimating the overall effect of extended and imagined contact we find that our estimate remains at d = 0.37, SE = 0.03. When we limit the analysis to studies of extended and imagined contact with $\geq$ 78 participants in the treatment condition, our findings also remain essentially the same: from d = 0.12, SE = 0.04 to d = 0.11, SE = 0.03.

Next we turn to Lai et al. (2014, 2016) who conducted two large-scale replication papers concerning interventions to change implicit prejudice. We find

that including these two papers does not materially affect our central estimates: d declines from 0.357 to 0.352, the standard error from 0.0209 to 0.0207.

Our overall results also remain stable when we include both the Many Labs results and the Lai (2014, 2016) results (d = 0.351, SE = 0.021), as well as when we differentiate results by outcome type. Note that this estimation procedure implicitly collapses all treatment arms within each study into a single estimated effect, rather than treating each intervention condition as its own independent estimate, and is thus conservative with respect to how much weight it accords any individual manuscript (see Appendix Section 2.1 for more details).

Lai et al. (2014, 2016) also provide many additional simultaneous measurements of explicit and implicit prejudice. When these new data points are included, the overall correlation between implicit and explicit outcomes remains minimal: changing from r=.02 without the Lai et al. studies to .04 when they are included.

3 List of Conversion Rules

let λ = the un-standardized effect size.
let $\Delta = \mu_t - \mu_c$
let DiD = $(\mu_{pre_t} - \mu_{post_t})$ - $(\mu_{pre_c} - \mu_{post_c})$

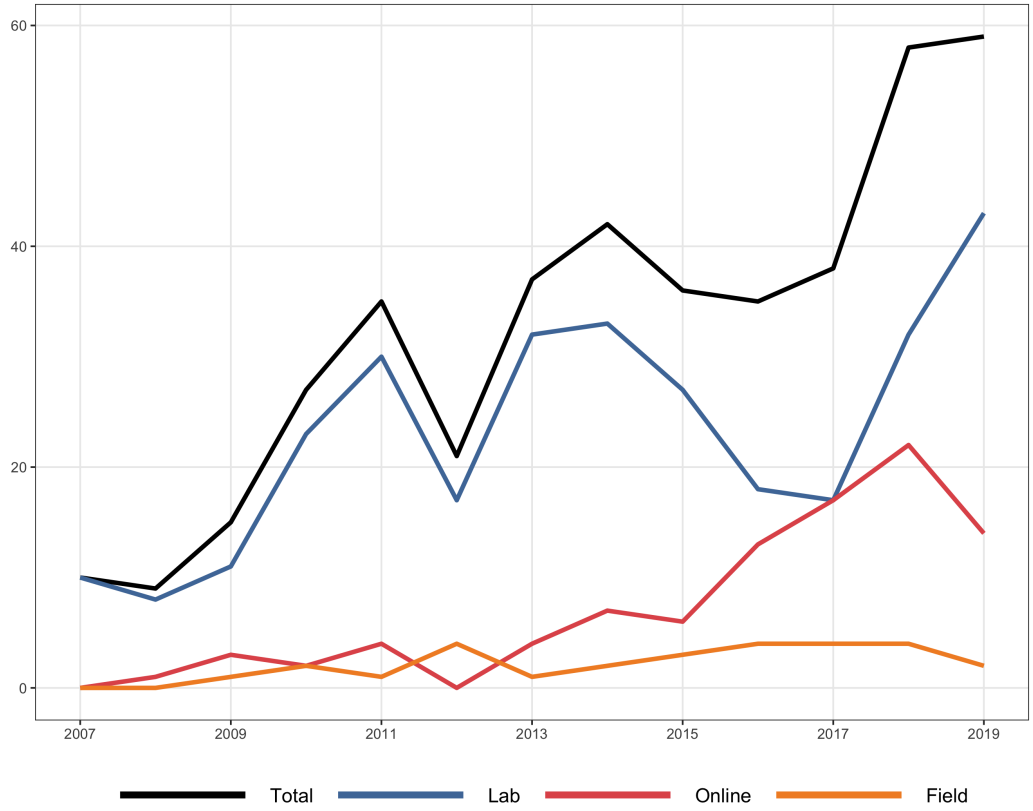
where t is the treatment condition, and c is the control condition
First d is calculated using from λ using the following function:

if Δ **then**
    $d = \frac{\lambda}{\sigma_c}$
else if d **then**
    $d = d$
else if η^2 **then**
    $d = 2 * \frac{\sqrt{\lambda}}{\sqrt{1-\lambda}}$
else if t **then**
    $d = \lambda * \sqrt{\frac{n_t+n_c}{n_t*n_c}}$
else if f **then**
    $d = \sqrt{\lambda * \frac{n_t+n_c}{n_t*n_c}}$
else if odds ratio **then**
    $d = ln\lambda * \sqrt{3}/\pi$
else if log odds ratio **then**
    $d = \lambda * \sqrt{3}/\pi$
else if β **then**
    $d = \frac{\lambda}{\sigma_{control}}$
else if Difference in proportions **then**
    $d = \frac{\lambda}{\sigma_{control}}$

The direction of the effect is determined by taking the absolute value of the d, and multiplying it by the sign of whether the effect reduced prejudice or not as understood by the coder's interpretation of the text.

$$|d| * sgn(\text{effect direction}) \tag{1}$$

4 Figure A: Time trends in prejudice reduction experiments



Type of study methods in past decade. This figure depicts the overall rise of experiments on prejudice reduction, and breaks down the total number of studies by those conducted in the laboratory, online, and in the field.

5 Studies by Quintile Group

Appendix Table 2: Lab Studies by Quintile Group

Sample Size	Number of Studies	Number of Articles	Effect Size	Standard Error
≤ 25	71	60	0.630	0.051
26 - 34	61	54	0.409	0.055
35 - 48	59	53	0.403	0.049
49 - 77	46	44	0.369	0.058
≥ 78	29	24	0.233	0.048

Appendix Table 3: Online Studies by Quintile Group

Sample Size	Number of Studies	Number of Articles	Effect Size	Standard Error
≤ 25	4	4	0.195	0.322
26 - 34	12	8	0.283	0.111
35 - 48	15	14	0.261	0.085
49 - 77	22	20	0.279	0.042
≥ 78	38	32	0.143	0.026

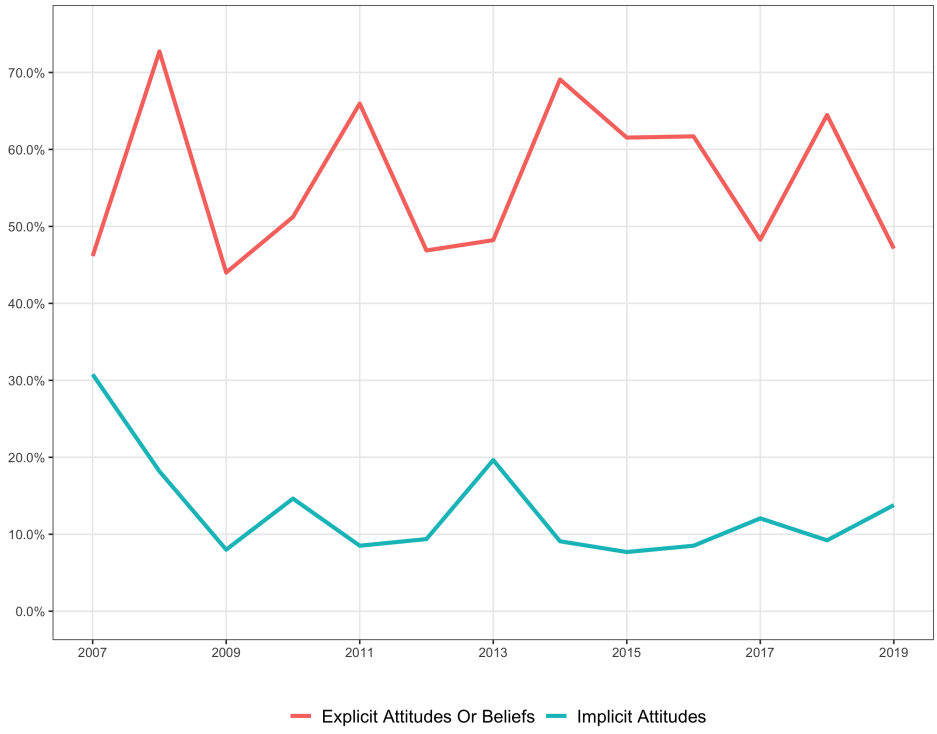
Appendix Table 4: Light Touch Studies by Quintile Group

Sample Size	Number of Studies	Number of Articles	Effect Size	Standard Error
≤ 25	63	54	0.599	0.060
26 - 34	65	55	0.372	0.045
35 - 48	63	56	0.380	0.045
49 - 77	61	57	0.344	0.046
≥ 78	56	44	0.165	0.024

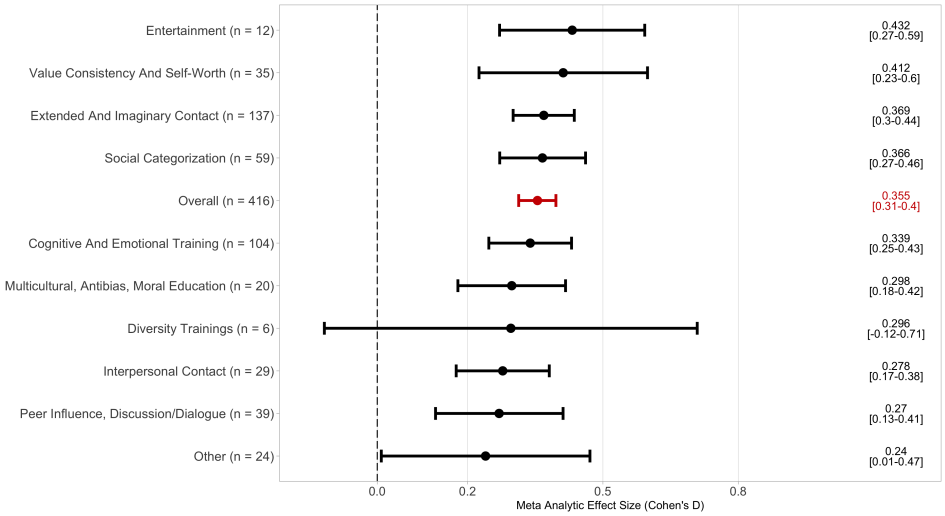
Appendix Table 5: Non-Light Touch Studies by Quintile Group

Sample Size	Number of Studies	Number of Articles	Effect Size	Standard Error
≤ 25	12	9	0.645	0.090
26 - 34	9	8	1.008	0.511
35 - 48	13	12	0.325	0.117
49 - 77	11	10	0.324	0.037
≥ 78	17	17	0.262	0.043

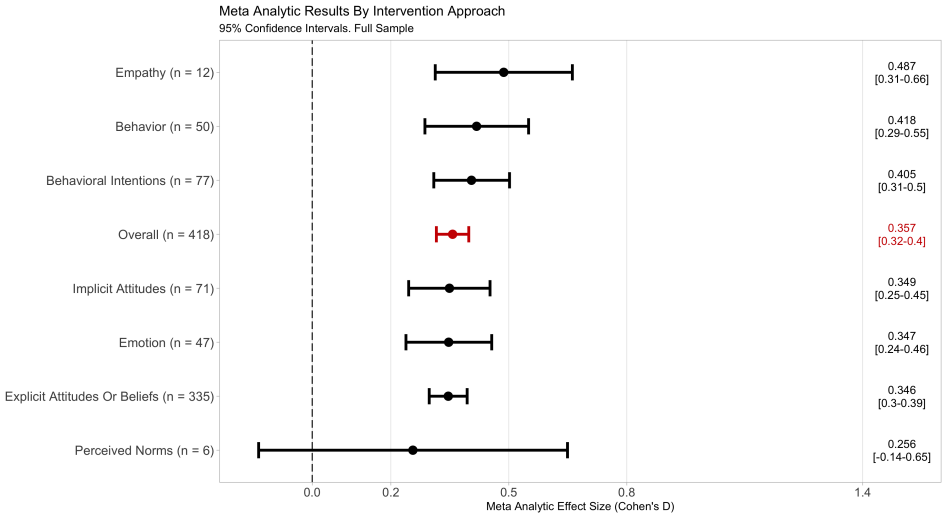
6 Additional Figures



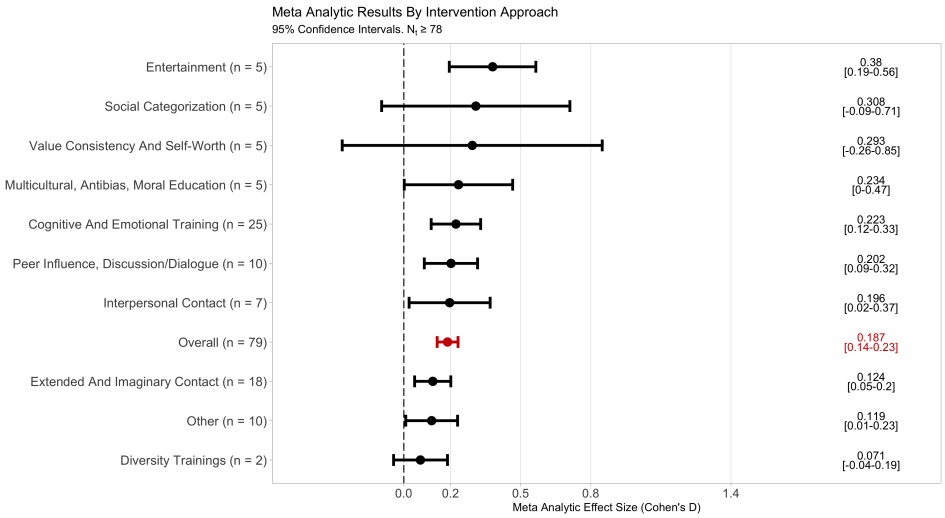
Appendix Figure 6: **Explicit and implicit attitudes over time as a percentage of yearly outcomes**



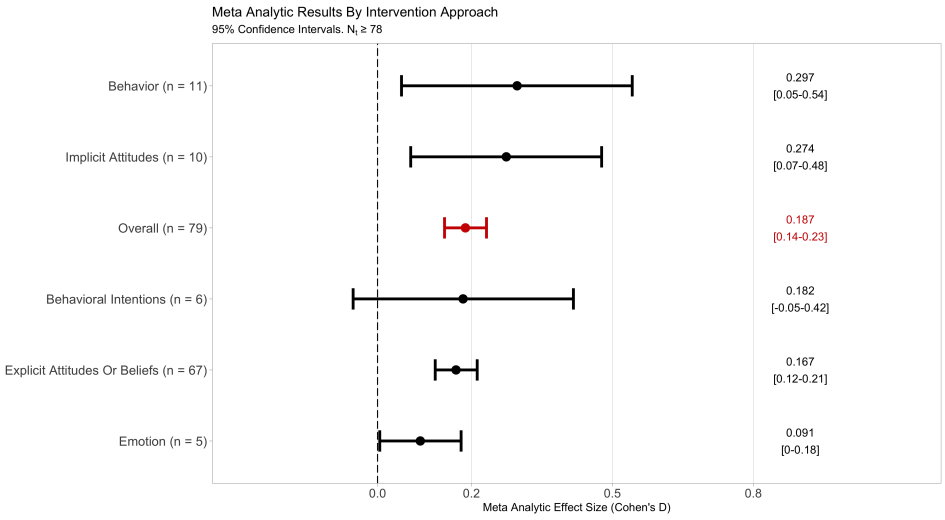
Appendix Figure 7: **Intervention approach meta-analytic estimates**



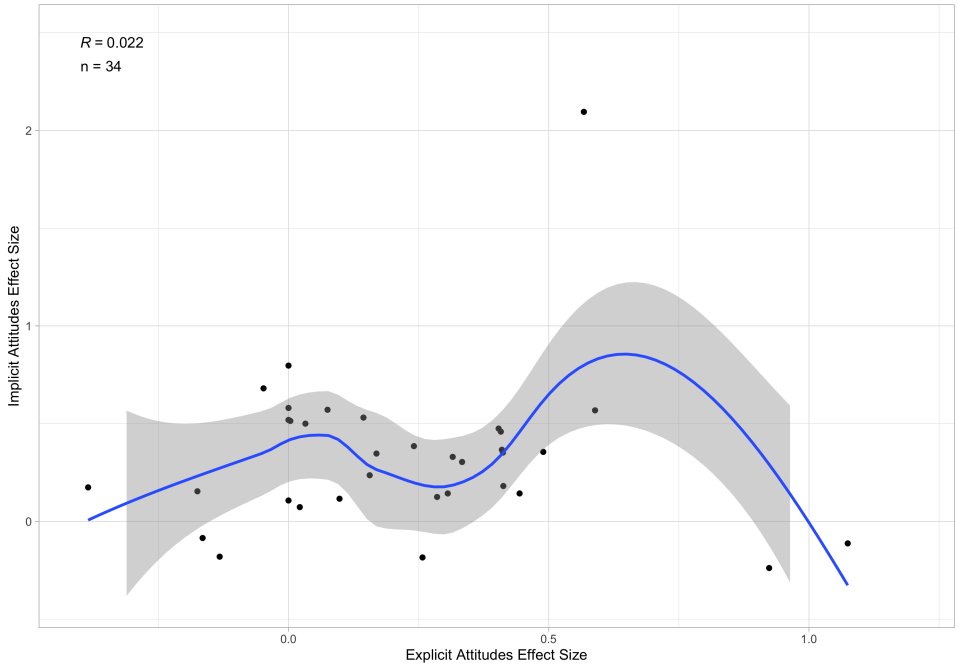
Appendix Figure 8: **Outcome type meta-analytic estimates**



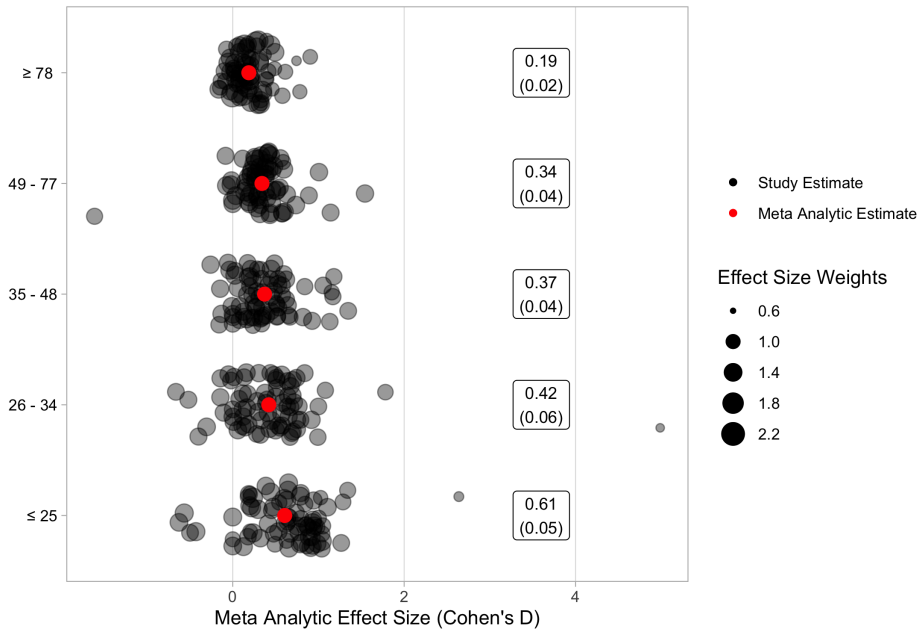
Appendix Figure 9: **Intervention approach meta-analytic estimates, for $\geq$ 78 participants in the treatment condition, no clusters**



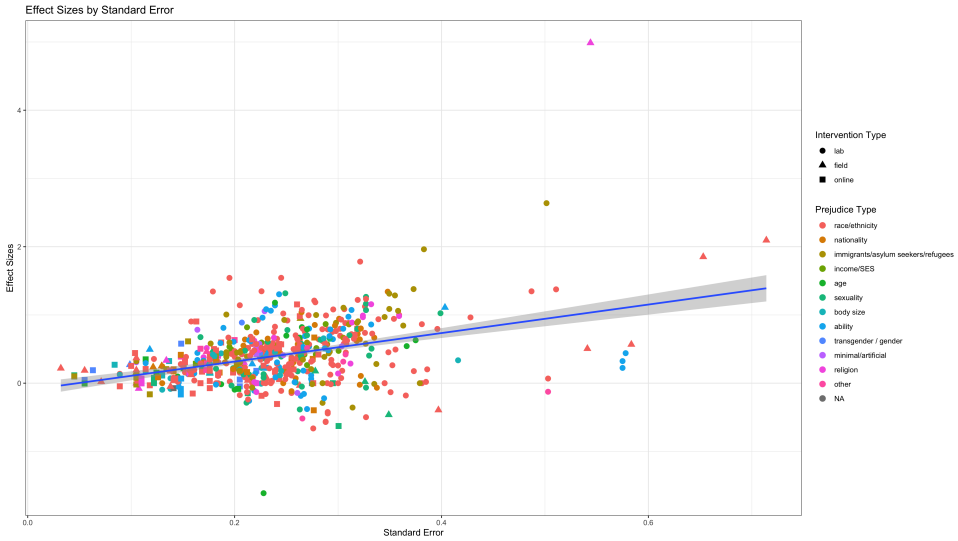
Appendix Figure 10: **Outcome types meta-analytic estimates, for $\geq$ 78 participants in the treatment condition, no clusters**



Appendix Figure 11: **Correspondence between explicit and implicit attitudes measured within the same study**



Appendix Figure 12: **Meta-analytic results by sample size quintiles** Boxed entries are meta-analytic results and standard errors, by subgroup. Effects collapsed on study level.



Appendix Figure 13: **Relationship between standard error and effect size, with point colors reflecting prejudice type and point shapes reflecting intervention type.**

7 Additional Sidebars

7.1 Audit Experiments to Measure Discrimination

For decades, researchers have assessed the extent of discrimination by employers, landlords, teachers, and public officials by examining how they react when encountering people of varying ethnicity, gender, or sexual orientation. For example, to measure housing discrimination, a researcher instructs a black confederate to inquire about the availability of an apartment that was listed for rent; an otherwise similar white confederate might inquire about the availability of the same apartment (Fang, Guess, and Humphreys 2019). If the rental agent informs the black couple that the apartment is unavailable but later tells the white couple that it is available, that is considered clear evidence of racial discrimination. This research method is now widely used to study discrimination in online interactions, such as job applications (Bertrand and Mullainathan 2004) or email inquiries to elected officials (Butler and Broockman 2011). Although this line of research is valuable for measuring the extent of discrimination as well as trends over time (Quillian et al. 2017), it has seldom been used as an outcome measure in an experiment designed to reduce prejudice. One exception is Chang et al. (2019, p.7780), which uses an audit experiment to assess the effects of diversity training, comparing rates of discrimination among those assigned to receive training versus those in the control group.

Bertrand, Marianne, and Sendhil Mullainathan. "Are Emily and Greg more

employable than Lakisha and Jamal? A field experiment on labor market discrimination." American economic review 94.4 (2004): 991-1013.
Butler, Daniel M., and David E. Broockman. "Do politicians racially discriminate against constituents? A field experiment on state legislators." American Journal of Political Science55.3 (2011): 463-477.
Fang, Albert H., Andrew M. Guess, and Macartan Humphreys. "Can the government deter discrimination? Evidence from a randomized intervention in New York City." The Journal of Politics 81.1 (2019): 127-141.
Quillian, Lincoln, Devah Pager, Ole Hexel, and Arnfinn H. Midtbøen. "Meta-analysis of field experiments shows no change in racial discrimination in hiring over time." Proceedings of the National Academy of Sciences 114, no. 41 (2017): 10870-10875.

7.2 Personalizing Prejudice Reduction Interventions

There is growing scholarly interest in customizing interventions according to the particular characteristics of the participants and the context they inhabit. Tests of personalized interventions look for evidence that treatments work especially well for certain kinds of people and in certain contexts. The search for especially strong treatment effects is of both practical and theoretical importance. If we had robust answers to this question, prejudice-reducing strategies could be targeted to specific people and specific contexts, akin to personalized medicine. Our collection of studies includes 50 that look for heterogeneous treatment effects by subjects' attributes or contextual characteristics. One concern is that a post hoc search for heterogeneous effects is prone to false discovery — with enough searching, evidence of significantly heterogeneous effects may turn up by chance. For this reason, searches of this kind are best done using automated machine-learning methods or following a pre-analysis plan. However, only three of the studies we reviewed took this kind of structured approach to the search for heterogeneous effects. Unplanned or exploratory investigation of treatment-by-covariate interactions are not necessarily misleading, but readers are advised to read personalization findings with caution until they are replicated with larger studies whose subgroup analyses are preregistered.

8 Acknowledgements

For excellent research assistance, we acknowledge Seth Green, Janet Pauketat, Ye Zhang, Siyeona Chang, Bryn McCarthy, Weiyu Wang, Gokul Ramachandran, Caroline Jones, Zara Riaz, Jackie Patterson, and Megan Testerman. Thanks to Calvin Lai, Tony Greenwald, Niv Reggev, and Eran Halperin for thoughtful feedback. This project would not have been possible without the research assistance of John Henry Pezzuto, and funding from Princeton University and the Rita Allen Foundation (to ELP) and from the Columbia University Graduate School of Arts and Sciences to DPG.

9 Dataverse link

Dataverse site for this study.

References

Abelson, R. P., & Prentice, D. A. (1989). Beliefs as possessions: A functional perspective. *Attitude structure and function*, *3*, 361–381.

Alimo, C. J. (2012). From dialogue to action: The impact of cross-race intergroup dialogue on the development of white college students as racial allies. *Equity & Excellence in Education*, *45*(1), 36–59.

Allport, G. (1954). The nature of prejudice.

Balas, R., & Sweklej, J. (2013). Changing prejudice with evaluative conditioning. *Polish Psychological Bulletin*, *44*(4), 379–383.

Banakou, D., Hanumanthu, P. D., & Slater, M. (2016). Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Frontiers in human neuroscience*, *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. *Journal of Experimental Social Psychology*, *49*(6), 1034–1041.

Bezrukova, K., Spell, C. S., Perry, J. L., & Jehn, K. A. (2016). A meta-analytical integration of over 40 years of research on diversity training evaluation. *Psychological Bulletin*, *142*(11), 1227.

Bigler, R. S. (1999). The use of multicultural curricula and materials to counter racism in children. *Journal of Social Issues*, *55*(4), 687–705.

Boag, E. M., & Wilson, D. (2014). Inside experience: Engagement empathy and prejudice towards prisoners. *Journal of Criminal Psychology*.

Brauer, M., & Er-Rafiy, A. (2011). Increasing perceived variability reduces prejudice and discrimination. *Journal of Experimental Social Psychology*, *47*(5), 871–881.

Broockman, D., & Kalla, J. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, *352*(6282), 220–224.

Brownstein, M., Madva, A., & Gawronski, B. (2019). What do implicit measures measure? *Wiley Interdisciplinary Reviews: Cognitive Science*, *10*(5), e1501.

Bruneau, E. G., & Saxe, R. (2012). The power of being heard: The benefits of 'perspective-giving'in the context of intergroup conflict. *Journal of experimental social psychology*, *48*(4), 855–866.

Cameron, L., Rutland, A., & Brown, R. (2007). Promoting children's positive intergroup attitudes towards stigmatized groups: Extended contact and multiple classification skills training. *International Journal of Behavioral Development*, *31*(5), 454–466.

Castillo, J. L. Á., Equizábal, A. J., Cámara, C. P., & González, H. G. (2014). The fight against prejudice in older adults: perspective taking effectiveness. *Revista Latinoamericana de Psicología*, *46*(3), 137–147.

Chaney, K. E. (2016). *Ends: the endurance, depth, and scope of confronting as a prejudice reduction strategy* (Unpublished doctoral dissertation). Rutgers University-Graduate School-New Brunswick.

Chang, E. H., Milkman, K. L., Gromet, D. M., Rebele, R. W., Massey, C., Duckworth, A. L., & Grant, A. M. (2019). The mixed effects of online diversity training. *Proceedings of the National Academy of Sciences*, *116*(16), 7778–7783.

Chongruksa, D., Prinyapol, P., Wadeng, Y., & Padungpong, C. (2010). Storytelling: program for multicultural understanding and respect among thai-buddhist and thai-muslim students. , *5*, 282–288. doi: 10.1016/j.sbspro.2010.07.089

Daumeyer, N. M., Onyeador, I. N., Brown, X., & Richeson, J. A. (2019, sep). Consequences of attributing discrimination to implicit vs. explicit bias. *Journal of Experimental Social Psychology*, *84*, 103812. Retrieved from `https://doi.org/10.1016%2Fj.jesp.2019.04.010` doi: 10.1016/j.jesp.2019.04.010

Dessel, A. B. (2010). Effects of intergroup dialogue: Public school teachers and sexual orientation prejudice. *Small Group Research*, *41*(5), 556–592.

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of experimental social psychology*, *48*(6), 1267–1278.

Devine, P. G., & Monteith, M. J. (1999). Automaticity and control in stereotyping.

Eno, C. A., & Ewoldsen, D. R. (2010). The influence of explicitly and implicitly measured prejudice on interpretations of and reactions to black film. *Media Psychology*, *13*(1), 1–30.

Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2). Stanford university press.

Finseraas, H., & Kotsadam, A. (2017). Does personal contact with ethnic minorities affect anti-immigrant sentiments? evidence from a field experiment. *European Journal of Political Research*, *56*(3), 703–722.

Fiske, S., & Lin, M. (81). Neuberg, s. l.(1999). the continuum model: Ten years later. *Dual Process theories in social psychology. New York: Guilford*.

Forscher, P. S., & Devine, P. G. (2017). Knowledge-based interventions are more likely to reduce legal disparities than are implicit bias interventions. Retrieved from https://doi.org/10.31234%2Fosf.io%2F8cgg5 doi: 10.31234/osf.io/8cgg5

Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2019, sep). A meta-analysis of procedures to change implicit measures. *Journal of Personality and Social Psychology*, *117*(3), 522–559. Retrieved from https://doi.org/10.1037%2Fpspa0000160 doi: 10.1037/pspa0000160

Forscher, P. S., Mitamura, C., Dix, E. L., Cox, W. T., & Devine, P. G. (2017). Breaking the prejudice habit: Mechanisms, timecourse, and longevity. *Journal of experimental social psychology*, *72*, 133–146.

French, A. R., Franz, T. M., Phelan, L. L., & Blaine, B. E. (2013). Reducing muslim/arab stereotypes through evaluative conditioning. *The Journal of social psychology*, *153*(1), 6–9.

Gaertner, S. L., Dovidio, J. F., Samuel, G., et al. (2000). *Reducing intergroup bias: The common ingroup identity model*. Psychology Press.

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of personality and social psychology*, *78*(4), 708.

Gelman, A. (2018, 3 15). *You need 16 times the sample size to estimate an interaction than to estimate a main effect Statistical Modeling, Causal Inference, and Social Science.* https://statmodeling.stat.columbia.edu/2018/03/15/need-16-times-sample-size-estimate-interaction-estimate-main-effect/.

Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton.

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of consumer Research*, *35*(3), 472–482.

Gómez, Á., Tropp, L. R., Vázquez, A., Voci, A., & Hewstone, M. (2018). Depersonalized extended contact and injunctive norms about cross-group friendship impact intergroup orientations. *Journal of Experimental Social Psychology*, *76*, 356–370.

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*(5), 701–721. Retrieved from https://doi.org/10.1037%2F0022-3514.79.5.701 doi: 10.1037/0022-3514.79.5.701

Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California Law Review*, *94*(4), 945–967.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. an improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. Retrieved from https://doi.org/10.1037%2F0022-3514.85.2.197 doi: 10.1037/0022-3514.85.2.197

Greitemeyer, T., & Schwab, A. (2014). Employing music exposure to reduce

prejudice and discrimination. *Aggressive Behavior*, *40*(6), 542–551.
Gross, J. J. (2013). *Handbook of emotion regulation*. Guilford publications.
Gurin, P., Nagda, B. R. A., & Zuniga, X. (2013). *Dialogue across difference: Practice, theory, and research on intergroup dialogue*. Russell Sage Foundation.
Hall, N. R., Crisp, R. J., & Suen, M.-w. (2009). Reducing implicit prejudice by blurring intergroup boundaries. *Basic and Applied Social Psychology*, *31*(3), 244–254.
Halperin, E., Porat, R., Tamir, M., & Gross, J. J. (2013). Can emotion regulation change political attitudes in intractable conflicts? from the laboratory to the field. *Psychological science*, *24*(1), 106–111.
Harper, J., & Carels, R. A. (2014). Impact of social pressure on stereotypes about obese people. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, *19*(3), 355–361.
Heider, F. (1958). *The psychology of interpersonal relations*. Psychology Press.
Heitland, K., & Bohner, G. (2010). Reducing prejudice via cognitive dissonance: Individual differences in preference for consistency moderate the effects of counter-attitudinal advocacy. *Social Influence*, *5*(3), 164–181.
Hocevar, K. P., Metzger, M., & Flanagin, A. J. (2017). Source credibility, expertise, and trust in health and risk messaging. In *Oxford research encyclopedia of communication*. Chicago.
Husnu, S., & Crisp, R. J. (2010a). Elaboration enhances the imagined contact effect. *Journal of Experimental Social Psychology*, *46*(6), 943–950.
Husnu, S., & Crisp, R. J. (2010b). Imagined intergroup contact: A new technique for encouraging greater inter-ethnic contact in cyprus. *Peace and Conflict: Journal of Peace Psychology*, *16*(1), 97–108. Retrieved from `https://doi.org/10.1080%2F10781910903484776` doi: 10.1080/10781910903484776
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 Processes & Intergroup Relations*, *22*(4), 511–529.
Johnson, D. R., Jasper, D. M., Griffin, S., & Huffman, B. L. (2013). Reading narrative fiction reduces arab-muslim prejudice and offers a safe haven from intergroup anxiety. *Social Cognition*, *31*(5), 578–598.
Jost, J. T., Rudman, L. A., Blair, I. V., Carney, D. R., Dasgupta, N., Glaser, J., & Hardin, C. D. (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. *Research in Organizational Behavior*, *29*, 39–69. Retrieved from `https://doi.org/10.1016%2Fj.riob.2009.10.001` doi: 10.1016/j.riob.2009.10.001
Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? assessing the efficacy of corporate affirmative action and diversity policies. *American sociological review*, *71*(4), 589–617.
Kalla, J. L., & Broockman, D. E. (2020). Reducing exclusionary attitudes through interpersonal conversation: evidence from three field experiments.

American Political Science Review, 1–16.

Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (close) distance makes the heart grow fonder: Improving implicit racial attitudes and interracial interactions through approach behaviors. *Journal of personality and social psychology*, *92*(6), 957.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., ... Nosek, B. A. (2014, may). Investigating variation in replicability: A "many labs" replication project. *Social Psychology*, *45*(3), 142–152. Retrieved from `https://doi.org/10.1027%2F1864-9335%2Fa000178` doi: 10.1027/1864-9335/a000178

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? *Annual Review of Organizational Psychology and Organizational Behavior*, *3*, 83–109.

Kurdi, B., Seitchik, A. E., Axt, J., Carroll, T., Karapetyan, A., Kaushik, N., ... Banaji, M. R. (2018, jun). Relationship between the implicit association test and intergroup behavior: A meta-analysis. Retrieved from `https://doi.org/10.31234%2Fosf.io%2F582gh` doi: 10.31234/osf.io/582gh

Lai, C., Hoffman, K., & Nosek, B. (2013, may). Reducing implicit prejudice. *Social and Personality Psychology Compass*, *7*(5), 315–330. Retrieved from `https://doi.org/10.1111%2Fspc3.12023` doi: 10.1111/spc3.12023

Lai, C., Marini, M., Lehr, S., Cerruti, C., Shin, J.-E., Joy-Gaba, J., ... others (2014). Reducing implicit racial preferences: I. a comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*(4), 1765. doi: 10.1037/a0036260

Lai, C., Skinner, A., Cooley, E., Murrar, S., Brauer, M., Devos, T., ... others (2016). Reducing implicit racial preferences: Ii. intervention effectiveness across time. *Journal of Experimental Psychology: General*, *145*(8), 1001-1016. doi: 10.1037/xge0000179

LaPiere, R. T. (1934, dec). Attitudes vs. actions. *Social Forces*, *13*(2), 230–237. Retrieved from `https://doi.org/10.2307%2F2570339` doi: 10.2307/2570339

Latu, I. M. (2010). Reducing automatic stereotype activation: Mechanisms and moderators of situational attribution training.

Lehmiller, J. J., Law, A. T., & Tormala, T. T. (2010). The effect of self-affirmation on sexual prejudice. *Journal of Experimental Social Psychology*, *46*(2), 276–285.

Lemmer, G., & Wagner, U. (2015). Can we really reduce ethnic prejudice outside the lab? a meta-analysis of direct and indirect contact interventions. *European Journal of Social Psychology*, *45*(2), 152–168.

Li, J., Fan, Y., Zhong, H.-Q., Duan, X.-L., Chen, W., Evans-Lacko, S., & Thornicroft, G. (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. *International journal of mental health systems*, *13*(1), 1.

Lowe, M. (2020). Types of contact: A field experiment on collaborative and

adversarial caste integration.

Lytle, A., & Levy, S. R. (2019). Reducing ageism: Education about aging and extended contact with older adults. *The Gerontologist*, *59*(3), 580–588.

Mendoza, S. A., Gollwitzer, P. M., & Amodio, D. M. (2010). Reducing the expression of implicit stereotypes: Reflexive control through implementation intentions. *Personality and Social Psychology Bulletin*, *36*(4), 512–523.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *Annals of internal medicine*, *151*(4), 264–269.

Mousa, S. (in press). Building tolerance: Intergroup contact and soccer in post-isis iraq.

Munger, K. (2017). Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior*, *39*(3), 629–649.

Murrar, S., & Brauer, M. (2018). Entertainment-education effectively reduces prejudice. *Group Processes & Intergroup Relations*, *21*(7), 1053–1077.

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. *Psychology of Music*, *44*(3), 388–398.

Oh, S. Y., Bailenson, J., Weisz, E., & Zaki, J. (2016). Virtually old: Embodied perspective taking and the reduction of ageism under threat. *Computers in Human Behavior*, *60*, 398–410.

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin*, *32*(4), 421–433.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013, aug). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*(2), 171–192. Retrieved from `https://doi.org/10.1037%2Fa0032734` doi: 10.1037/a0032734

Paluck, E. L. (2006). Diversity training and intergroup contact: A call to action research. *Journal of Social Issues*, *62*(3), 577–595.

Paluck, E. L. (2009a). Reducing intergroup prejudice and conflict using the media: a field experiment in rwanda. *Journal of personality and social psychology*, *96*(3), 574.

Paluck, E. L. (2009b). What's in a norm? sources and processes of norm change.

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? a review and assessment of research and practice. *Annual review of psychology*, *60*, 339–367.

Paluck, E. L., Green, S. A., & Green, D. P. (2019). The contact hypothesis re-evaluated. *Behavioural Public Policy*, *3*(2), 129–158.

Paluck, E. L., & Prentice, D. (2020). The social norms revolution will not be individualized: Beliefs about social norms are shaped by group experience. *Personality and Social Psychology Bulletin*.

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. *Journal of Personality and Social Psychology*,

42(6), 1051.

Parrott, S., Carpentier, F. R. D., & Northup, C. T. (2017). A test of interactive narrative as a tool against prejudice. *Howard Journal of Communications*, *28*(4), 374–389.

Patel, S. L. (2013). *Examining the influence of perceived social consensus information on weight prejudice across development*. The University of Texas at Dallas.

Perry, S. (2011). Responses to prejudice feedback on the race-implicit associations test and the role of bias awareness [dissertation]. *Chicago (IL): University of Illinois at Chicago*.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of personality and social psychology*, *90*(5), 751.

Radnitz, S. (2018). Historical narratives and post-conflict reconciliation: An experiment in azerbaijan. *Conflict Management and Peace Science*, *35*(2), 154–174.

Robinson, C. (2010). Cross-cutting messages and political tolerance: An experiment using evangelical protestants. *Political Behavior*, *32*(4), 495–515.

Roese, N. J., & Jamieson, D. W. (1993). Twenty years of bogus pipeline research: A critical review and meta-analysis. *Psychological Bulletin*, *114*(2), 363–375. Retrieved from `https://doi.org/10.1037%2F0033-2909.114.2.363` doi: 10.1037/0033-2909.114.2.363

Rokeach, M. (1971). Long-range experimental modification of values, attitudes, and behavior. *American psychologist*, *26*(5), 453.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological bulletin*, *86*(3), 638.

Scacco, A., & Warren, S. S. (2018). Can social contact reduce prejudice and discrimination? evidence from a field experiment in nigeria. *American Political Science Review*, *112*(3), 654–677.

Sechrist, G. B., & Milford-Szafran, L. R. (2011). "i depend on you, you depend on me. shouldn't we agree?": The influence of interdependent relationships on individuals' racial attitudes. *Basic and applied social psychology*, *33*(2), 145–156.

Sekaquaptewa, D., Espinoza, P., Thompson, M., Vargas, P., & von Hippel, W. (2003). Stereotypic explanatory bias: Implicit stereotyping as a predictor of discrimination. *Journal of Experimental Social Psychology*, *39*(1), 75–82.

Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. *Advances in experimental social psychology*, *38*, 183–242.

Shih, M. J., Stotzer, R., & Gutiérrez, A. S. (2013). Perspective-taking and empathy: Generalizing the reduction of group bias towards asian americans to general outgroups. *Asian American Journal of Psychology*, *4*(2), 79.

Siegel, A. A., & Badaan, V. (in press). no2sectarianism: Experimental approaches to reducing sectarian hate speech online.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of experimental psychology: General*, *143*(2), 534.

Smeekes, A., Verkuyten, M., & Poppe, E. (2012). How a tolerant past affects the present: Historical tolerance and the acceptance of muslim expressive rights. *Personality and Social Psychology Bulletin*, *38*(11), 1410–1422.

Smith, B. D., & Silk, K. (2011). Cultural competence clinic: an online, interactive, simulation for working effectively with arab american muslim patients. *Academic Psychiatry*, *35*(5), 312–316.

Stell, A. J., & Farsides, T. (2016). Brief loving-kindness meditation reduces racial bias, mediated by positive other-regarding emotions. *Motivation and Emotion*, *40*(1), 140–147.

Stewart, B. D., & Payne, B. K. (2008). Bringing automatic stereotyping under control: Implementation intentions as efficient means of thought control. *Personality and Social Psychology Bulletin*, *34*(10), 1332–1345.

Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. in wg austin & s. worchel (eds.), the social psychology of intergroup relations (pp. 33-47). *Monterey, CA: Brooks/Cole*.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Turner, R. N., Crisp, R. J., & Lambert, E. (2007). Imagining intergroup contact can improve intergroup attitudes. *Group Processes & Intergroup Relations*, *10*(4), 427–441.

Ugarriza, J. E., & Nussio, E. (2017). The effect of perspective-giving on postconflict reconciliation. an experimental approach. *Political Psychology*, *38*(1), 3–19.

Vezzali, L. (2017). Valence matters: Positive meta-stereotypes and interethnic interactions. *The Journal of social psychology*, *157*(2), 247–261.

Vezzali, L., Stathi, S., Crisp, R. J., & Capozza, D. (2015). Comparing direct and imagined intergroup contact among children: Effects on outgroup stereotypes and helping intentions. *International Journal of Intercultural Relations*, *49*, 46–53.

Villicana, A. J., Rivera, L. M., & Garcia, D. M. (2018). When one's group is beneficial: The effect of group-affirmation and subjective group identification on prejudice. *Group Processes & Intergroup Relations*, *21*(6), 962–976.

Walsh, S. P. (2013). *Reducing automatic stereotype activation: European-and african american photos in situational attribution training* (Unpublished doctoral dissertation). University of Mississippi.

Walton, G. M. (2014). The new science of wise psychological interventions. *Current Directions in Psychological Science*, *23*(1), 73–82.

West, K., & Bruckmüller, S. (2013). Nice and easy does it: How perceptual fluency moderates the effectiveness of imagined contact. *Journal of Experimental Social Psychology*, *49*(2), 254–262.

Witkowska, M., Beneda, M., Čehajić-Clancy, S., & Bilewicz, M. (2019). Fostering contact after historical atrocities: The potential of moral exemplars. *Political Psychology*, *40*(3), 565–582.

Woodcock, A., & Monteith, M. J. (2013). Forging links with the self to combat implicit bias. *Group processes & intergroup relations*, *16*(4), 445–461.

Wright, S. C., Aron, A., McLaughlin-Volpe, T., & Ropp, S. A. (1997). The

extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social psychology*, *73*(1), 73.

Yablon, Y. B. (2012). Are we preaching to the converted? the role of motivation in understanding the contribution of intergroup encounters. *Journal of Peace Education*, *9*(3), 249–263.

10 Studies Used in Meta-Analysis

References

Abelson, R. P., & Prentice, D. A. (1989). Beliefs as possessions: A functional perspective. *Attitude structure and function*, *3*, 361–381.

Alimo, C. J. (2012). From dialogue to action: The impact of cross-race intergroup dialogue on the development of white college students as racial allies. *Equity & Excellence in Education*, *45*(1), 36–59.

Allport, G. (1954). The nature of prejudice.

Balas, R., & Sweklej, J. (2013). Changing prejudice with evaluative conditioning. *Polish Psychological Bulletin*, *44*(4), 379–383.

Banakou, D., Hanumanthu, P. D., & Slater, M. (2016). Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Frontiers in human neuroscience*, *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. *Journal of Experimental Social Psychology*, *49*(6), 1034–1041.

Bezrukova, K., Spell, C. S., Perry, J. L., & Jehn, K. A. (2016). A meta-analytical integration of over 40 years of research on diversity training evaluation. *Psychological Bulletin*, *142*(11), 1227.

Bigler, R. S. (1999). The use of multicultural curricula and materials to counter racism in children. *Journal of Social Issues*, *55*(4), 687–705.

Boag, E. M., & Wilson, D. (2014). Inside experience: Engagement empathy and prejudice towards prisoners. *Journal of Criminal Psychology*.

Brauer, M., & Er-Rafiy, A. (2011). Increasing perceived variability reduces prejudice and discrimination. *Journal of Experimental Social Psychology*, *47*(5), 871–881.

Broockman, D., & Kalla, J. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, *352*(6282), 220–224.

Brownstein, M., Madva, A., & Gawronski, B. (2019). What do implicit measures measure? *Wiley Interdisciplinary Reviews: Cognitive Science*, *10*(5), e1501.

Bruneau, E. G., & Saxe, R. (2012). The power of being heard: The benefits of 'perspective-giving'in the context of intergroup conflict. *Journal of experimental social psychology*, *48*(4), 855–866.

Cameron, L., Rutland, A., & Brown, R. (2007). Promoting children's positive intergroup attitudes towards stigmatized groups: Extended contact and multiple classification skills training. *International Journal of Behavioral Development*, *31*(5), 454–466.

Castillo, J. L. Á., Equizábal, A. J., Cámara, C. P., & González, H. G. (2014). The fight against prejudice in older adults: perspective taking effectiveness. *Revista Latinoamericana de Psicología*, *46*(3), 137–147.

Chaney, K. E. (2016). *Ends: the endurance, depth, and scope of confronting as a prejudice reduction strategy* (Unpublished doctoral dissertation). Rutgers University-Graduate School-New Brunswick.

Chang, E. H., Milkman, K. L., Gromet, D. M., Rebele, R. W., Massey, C., Duckworth, A. L., & Grant, A. M. (2019). The mixed effects of online diversity training. *Proceedings of the National Academy of Sciences*, *116*(16), 7778–7783.

Chongruksa, D., Prinyapol, P., Wadeng, Y., & Padungpong, C. (2010). Storytelling: program for multicultural understanding and respect among thai-buddhist and thai-muslim students. , *5*, 282–288. doi: 10.1016/j.sbspro.2010.07.089

Daumeyer, N. M., Onyeador, I. N., Brown, X., & Richeson, J. A. (2019, sep). Consequences of attributing discrimination to implicit vs. explicit bias. *Journal of Experimental Social Psychology*, *84*, 103812. Retrieved from `https://doi.org/10.1016%2Fj.jesp.2019.04.010` doi: 10.1016/j.jesp.2019.04.010

Dessel, A. B. (2010). Effects of intergroup dialogue: Public school teachers and sexual orientation prejudice. *Small Group Research*, *41*(5), 556–592.

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of experimental social psychology*, *48*(6), 1267–1278.

Devine, P. G., & Monteith, M. J. (1999). Automaticity and control in stereotyping.

Eno, C. A., & Ewoldsen, D. R. (2010). The influence of explicitly and implicitly measured prejudice on interpretations of and reactions to black film. *Media Psychology*, *13*(1), 1–30.

Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2). Stanford university press.

Finseraas, H., & Kotsadam, A. (2017). Does personal contact with ethnic minorities affect anti-immigrant sentiments? evidence from a field experiment. *European Journal of Political Research*, *56*(3), 703–722.

Fiske, S., & Lin, M. (81). Neuberg, s. l.(1999). the continuum model: Ten years later. *Dual Process theories in social psychology. New York: Guilford*.

Forscher, P. S., & Devine, P. G. (2017). Knowledge-based interventions are more likely to reduce legal disparities than are implicit bias interventions. Retrieved from `https://doi.org/10.31234%2Fosf.io%2F8cgg5` doi: 10.31234/osf.io/8cgg5

Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2019, sep). A meta-analysis of procedures to change

implicit measures. *Journal of Personality and Social Psychology*, *117*(3), 522–559. Retrieved from https://doi.org/10.1037%2Fpspa0000160 doi: 10.1037/pspa0000160

Forscher, P. S., Mitamura, C., Dix, E. L., Cox, W. T., & Devine, P. G. (2017). Breaking the prejudice habit: Mechanisms, timecourse, and longevity. *Journal of experimental social psychology*, *72*, 133–146.

French, A. R., Franz, T. M., Phelan, L. L., & Blaine, B. E. (2013). Reducing muslim/arab stereotypes through evaluative conditioning. *The Journal of social psychology*, *153*(1), 6–9.

Gaertner, S. L., Dovidio, J. F., Samuel, G., et al. (2000). *Reducing intergroup bias: The common ingroup identity model*. Psychology Press.

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of personality and social psychology*, *78*(4), 708.

Gelman, A. (2018, 3 15). *You need 16 times the sample size to estimate an interaction than to estimate a main effect Statistical Modeling, Causal Inference, and Social Science.* https://statmodeling.stat.columbia.edu/2018/03/15/need-16-times-sample-size-estimate-interaction-estimate-main-effect/.

Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton.

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of consumer Research*, *35*(3), 472–482.

Gómez, Á., Tropp, L. R., Vázquez, A., Voci, A., & Hewstone, M. (2018). Depersonalized extended contact and injunctive norms about cross-group friendship impact intergroup orientations. *Journal of Experimental Social Psychology*, *76*, 356–370.

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*(5), 701–721. Retrieved from https://doi.org/10.1037%2F0022-3514.79.5.701 doi: 10.1037/0022-3514.79.5.701

Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California Law Review*, *94*(4), 945–967.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. an improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. Retrieved from https://doi.org/10.1037%2F0022-3514.85.2.197 doi: 10.1037/0022-3514.85.2.197

Greitemeyer, T., & Schwab, A. (2014). Employing music exposure to reduce prejudice and discrimination. *Aggressive Behavior*, *40*(6), 542–551.

Gross, J. J. (2013). *Handbook of emotion regulation*. Guilford publications.

Gurin, P., Nagda, B. R. A., & Zuniga, X. (2013). *Dialogue across difference: Practice, theory, and research on intergroup dialogue*. Russell Sage Foundation.

Hall, N. R., Crisp, R. J., & Suen, M.-w. (2009). Reducing implicit prejudice

by blurring intergroup boundaries. *Basic and Applied Social Psychology*, *31*(3), 244–254.

Halperin, E., Porat, R., Tamir, M., & Gross, J. J. (2013). Can emotion regulation change political attitudes in intractable conflicts? from the laboratory to the field. *Psychological science*, *24*(1), 106–111.

Harper, J., & Carels, R. A. (2014). Impact of social pressure on stereotypes about obese people. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, *19*(3), 355–361.

Heider, F. (1958). *The psychology of interpersonal relations*. Psychology Press.

Heitland, K., & Bohner, G. (2010). Reducing prejudice via cognitive dissonance: Individual differences in preference for consistency moderate the effects of counter-attitudinal advocacy. *Social Influence*, *5*(3), 164–181.

Hocevar, K. P., Metzger, M., & Flanagin, A. J. (2017). Source credibility, expertise, and trust in health and risk messaging. In *Oxford research encyclopedia of communication*. Chicago.

Husnu, S., & Crisp, R. J. (2010a). Elaboration enhances the imagined contact effect. *Journal of Experimental Social Psychology*, *46*(6), 943–950.

Husnu, S., & Crisp, R. J. (2010b). Imagined intergroup contact: A new technique for encouraging greater inter-ethnic contact in cyprus. *Peace and Conflict: Journal of Peace Psychology*, *16*(1), 97–108. Retrieved from `https://doi.org/10.1080%2F10781910903484776` doi: 10.1080/10781910903484776

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 Processes & Intergroup Relations*, *22*(4), 511–529.

Johnson, D. R., Jasper, D. M., Griffin, S., & Huffman, B. L. (2013). Reading narrative fiction reduces arab-muslim prejudice and offers a safe haven from intergroup anxiety. *Social Cognition*, *31*(5), 578–598.

Jost, J. T., Rudman, L. A., Blair, I. V., Carney, D. R., Dasgupta, N., Glaser, J., & Hardin, C. D. (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. *Research in Organizational Behavior*, *29*, 39–69. Retrieved from `https://doi.org/10.1016%2Fj.riob.2009.10.001` doi: 10.1016/j.riob.2009.10.001

Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? assessing the efficacy of corporate affirmative action and diversity policies. *American sociological review*, *71*(4), 589–617.

Kalla, J. L., & Broockman, D. E. (2020). Reducing exclusionary attitudes through interpersonal conversation: evidence from three field experiments. *American Political Science Review*, 1–16.

Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (close) distance makes the heart grow fonder: Improving implicit racial attitudes and interracial interactions through approach behaviors. *Journal of personality and social psychology*, *92*(6), 957.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein,

M. J., ... Nosek, B. A. (2014, may). Investigating variation in replicability: A "many labs" replication project. *Social Psychology*, *45*(3), 142–152. Retrieved from https://doi.org/10.1027%2F1864-9335%2Fa000178 doi: 10.1027/1864-9335/a000178

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? *Annual Review of Organizational Psychology and Organizational Behavior*, *3*, 83–109.

Kurdi, B., Seitchik, A. E., Axt, J., Carroll, T., Karapetyan, A., Kaushik, N., ... Banaji, M. R. (2018, jun). Relationship between the implicit association test and intergroup behavior: A meta-analysis. Retrieved from https://doi.org/10.31234%2Fosf.io%2F582gh doi: 10.31234/osf.io/582gh

Lai, C., Hoffman, K., & Nosek, B. (2013, may). Reducing implicit prejudice. *Social and Personality Psychology Compass*, *7*(5), 315–330. Retrieved from https://doi.org/10.1111%2Fspc3.12023 doi: 10.1111/spc3.12023

Lai, C., Marini, M., Lehr, S., Cerruti, C., Shin, J.-E., Joy-Gaba, J., ... others (2014). Reducing implicit racial preferences: I. a comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, *143*(4), 1765. doi: 10.1037/a0036260

Lai, C., Skinner, A., Cooley, E., Murrar, S., Brauer, M., Devos, T., ... others (2016). Reducing implicit racial preferences: Ii. intervention effectiveness across time. *Journal of Experimental Psychology: General*, *145*(8), 1001-1016. doi: 10.1037/xge0000179

LaPiere, R. T. (1934, dec). Attitudes vs. actions. *Social Forces*, *13*(2), 230–237. Retrieved from https://doi.org/10.2307%2F2570339 doi: 10.2307/2570339

Latu, I. M. (2010). Reducing automatic stereotype activation: Mechanisms and moderators of situational attribution training.

Lehmiller, J. J., Law, A. T., & Tormala, T. T. (2010). The effect of self-affirmation on sexual prejudice. *Journal of Experimental Social Psychology*, *46*(2), 276–285.

Lemmer, G., & Wagner, U. (2015). Can we really reduce ethnic prejudice outside the lab? a meta-analysis of direct and indirect contact interventions. *European Journal of Social Psychology*, *45*(2), 152–168.

Li, J., Fan, Y., Zhong, H.-Q., Duan, X.-L., Chen, W., Evans-Lacko, S., & Thornicroft, G. (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. *International journal of mental health systems*, *13*(1), 1.

Lowe, M. (2020). Types of contact: A field experiment on collaborative and adversarial caste integration.

Lytle, A., & Levy, S. R. (2019). Reducing ageism: Education about aging and extended contact with older adults. *The Gerontologist*, *59*(3), 580–588.

Mendoza, S. A., Gollwitzer, P. M., & Amodio, D. M. (2010). Reducing the expression of implicit stereotypes: Reflexive control through implementation intentions. *Personality and Social Psychology Bulletin*, *36*(4), 512–523.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *Annals of internal medicine*, *151*(4), 264–269.

Mousa, S. (in press). Building tolerance: Intergroup contact and soccer in post-isis iraq.

Munger, K. (2017). Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior*, *39*(3), 629–649.

Murrar, S., & Brauer, M. (2018). Entertainment-education effectively reduces prejudice. *Group Processes & Intergroup Relations*, *21*(7), 1053–1077.

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. *Psychology of Music*, *44*(3), 388–398.

Oh, S. Y., Bailenson, J., Weisz, E., & Zaki, J. (2016). Virtually old: Embodied perspective taking and the reduction of ageism under threat. *Computers in Human Behavior*, *60*, 398–410.

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin*, *32*(4), 421–433.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013, aug). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*(2), 171–192. Retrieved from `https://doi.org/10.1037%2Fa0032734` doi: 10.1037/a0032734

Paluck, E. L. (2006). Diversity training and intergroup contact: A call to action research. *Journal of Social Issues*, *62*(3), 577–595.

Paluck, E. L. (2009a). Reducing intergroup prejudice and conflict using the media: a field experiment in rwanda. *Journal of personality and social psychology*, *96*(3), 574.

Paluck, E. L. (2009b). What's in a norm? sources and processes of norm change.

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? a review and assessment of research and practice. *Annual review of psychology*, *60*, 339–367.

Paluck, E. L., Green, S. A., & Green, D. P. (2019). The contact hypothesis re-evaluated. *Behavioural Public Policy*, *3*(2), 129–158.

Paluck, E. L., & Prentice, D. (2020). The social norms revolution will not be individualized: Beliefs about social norms are shaped by group experience. *Personality and Social Psychology Bulletin*.

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. *Journal of Personality and Social Psychology*, *42*(6), 1051.

Parrott, S., Carpentier, F. R. D., & Northup, C. T. (2017). A test of interactive narrative as a tool against prejudice. *Howard Journal of Communications*, *28*(4), 374–389.

Patel, S. L. (2013). *Examining the influence of perceived social consensus information on weight prejudice across development*. The University of

Texas at Dallas.

Perry, S. (2011). Responses to prejudice feedback on the race-implicit associations test and the role of bias awareness [dissertation]. *Chicago (IL): University of Illinois at Chicago*.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of personality and social psychology*, *90*(5), 751.

Radnitz, S. (2018). Historical narratives and post-conflict reconciliation: An experiment in azerbaijan. *Conflict Management and Peace Science*, *35*(2), 154–174.

Robinson, C. (2010). Cross-cutting messages and political tolerance: An experiment using evangelical protestants. *Political Behavior*, *32*(4), 495–515.

Roese, N. J., & Jamieson, D. W. (1993). Twenty years of bogus pipeline research: A critical review and meta-analysis. *Psychological Bulletin*, *114*(2), 363–375. Retrieved from `https://doi.org/10.1037%2F0033-2909.114.2.363` doi: 10.1037/0033-2909.114.2.363

Rokeach, M. (1971). Long-range experimental modification of values, attitudes, and behavior. *American psychologist*, *26*(5), 453.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological bulletin*, *86*(3), 638.

Scacco, A., & Warren, S. S. (2018). Can social contact reduce prejudice and discrimination? evidence from a field experiment in nigeria. *American Political Science Review*, *112*(3), 654–677.

Sechrist, G. B., & Milford-Szafran, L. R. (2011). "i depend on you, you depend on me. shouldn't we agree?": The influence of interdependent relationships on individuals' racial attitudes. *Basic and applied social psychology*, *33*(2), 145–156.

Sekaquaptewa, D., Espinoza, P., Thompson, M., Vargas, P., & von Hippel, W. (2003). Stereotypic explanatory bias: Implicit stereotyping as a predictor of discrimination. *Journal of Experimental Social Psychology*, *39*(1), 75–82.

Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. *Advances in experimental social psychology*, *38*, 183–242.

Shih, M. J., Stotzer, R., & Gutiérrez, A. S. (2013). Perspective-taking and empathy: Generalizing the reduction of group bias towards asian americans to general outgroups. *Asian American Journal of Psychology*, *4*(2), 79.

Siegel, A. A., & Badaan, V. (in press). no2sectarianism: Experimental approaches to reducing sectarian hate speech online.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of experimental psychology: General*, *143*(2), 534.

Smeekes, A., Verkuyten, M., & Poppe, E. (2012). How a tolerant past affects the present: Historical tolerance and the acceptance of muslim expressive rights. *Personality and Social Psychology Bulletin*, *38*(11), 1410–1422.

Smith, B. D., & Silk, K. (2011). Cultural competence clinic: an online, interactive, simulation for working effectively with arab american muslim patients. *Academic Psychiatry*, *35*(5), 312–316.

Stell, A. J., & Farsides, T. (2016). Brief loving-kindness meditation reduces racial bias, mediated by positive other-regarding emotions. *Motivation and Emotion*, *40*(1), 140–147.

Stewart, B. D., & Payne, B. K. (2008). Bringing automatic stereotyping under control: Implementation intentions as efficient means of thought control. *Personality and Social Psychology Bulletin*, *34*(10), 1332–1345.

Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. in wg austin & s. worchel (eds.), the social psychology of intergroup relations (pp. 33-47). *Monterey, CA: Brooks/Cole*.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Turner, R. N., Crisp, R. J., & Lambert, E. (2007). Imagining intergroup contact can improve intergroup attitudes. *Group Processes & Intergroup Relations*, *10*(4), 427–441.

Ugarriza, J. E., & Nussio, E. (2017). The effect of perspective-giving on postconflict reconciliation. an experimental approach. *Political Psychology*, *38*(1), 3–19.

Vezzali, L. (2017). Valence matters: Positive meta-stereotypes and interethnic interactions. *The Journal of social psychology*, *157*(2), 247–261.

Vezzali, L., Stathi, S., Crisp, R. J., & Capozza, D. (2015). Comparing direct and imagined intergroup contact among children: Effects on outgroup stereotypes and helping intentions. *International Journal of Intercultural Relations*, *49*, 46–53.

Villicana, A. J., Rivera, L. M., & Garcia, D. M. (2018). When one's group is beneficial: The effect of group-affirmation and subjective group identification on prejudice. *Group Processes & Intergroup Relations*, *21*(6), 962–976.

Walsh, S. P. (2013). *Reducing automatic stereotype activation: European-and african american photos in situational attribution training* (Unpublished doctoral dissertation). University of Mississippi.

Walton, G. M. (2014). The new science of wise psychological interventions. *Current Directions in Psychological Science*, *23*(1), 73–82.

West, K., & Bruckmüller, S. (2013). Nice and easy does it: How perceptual fluency moderates the effectiveness of imagined contact. *Journal of Experimental Social Psychology*, *49*(2), 254–262.

Witkowska, M., Beneda, M., Čehajić-Clancy, S., & Bilewicz, M. (2019). Fostering contact after historical atrocities: The potential of moral exemplars. *Political Psychology*, *40*(3), 565–582.

Woodcock, A., & Monteith, M. J. (2013). Forging links with the self to combat implicit bias. *Group processes & intergroup relations*, *16*(4), 445–461.

Wright, S. C., Aron, A., McLaughlin-Volpe, T., & Ropp, S. A. (1997). The extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social psychology*, *73*(1), 73.

Yablon, Y. B. (2012). Are we preaching to the converted? the role of motivation in understanding the contribution of intergroup encounters. *Journal of Peace Education*, *9*(3), 249–263.