

What reduces prejudice in the real world? A meta-analysis of prejudice reduction field experiments

Author

Hsieh, Wing, Faulkner, Nicholas, Wickes, Rebecca

Published

2021

Journal Title

British Journal of Social Psychology

Version

Accepted Manuscript (AM)

DOI

[10.1111/bjso.12509](https://doi.org/10.1111/bjso.12509)

Downloaded from

<http://hdl.handle.net/10072/416086>

Griffith Research Online

<https://research-repository.griffith.edu.au>

What reduces prejudice in the real world? A meta-analysis of prejudice reduction field experiments

Wing Hsieh^{1*}, Nicholas Faulkner¹ and Rebecca Wickes²

¹ BehaviourWorks Australia, Monash Sustainable Development Institute, Monash University,
Melbourne, Australia

² School of Social Sciences, Monash University, Melbourne, Australia

* Corresponding author

Email: wing.hsieh@monash.edu

Pre-print: 26 August 2020

Abstract

Despite advances to promote acceptance and equity in many countries, prejudice remains a significant social problem. Promoting intergroup harmony requires knowledge about what works to reduce prejudice in community settings. This systematic review and meta-analysis of field experiments undertaken since 2009 reveals the most effective intervention types and conditions for reducing negative attitudes towards minority groups in real-world settings. Across 34 intervention arms and 7522 participants, results reveal that prejudice reduction interventions are effective at improving attitudes towards minority groups in real world settings ($SMD = 0.37$, [0.19; 0.56] 95 per cent confidence interval). Subgroup analyses indicate that some of the less commonly explored approaches, such as perceived variability and perspective taking, may have larger effect sizes than those using contact-based interventions. Additionally, results show that interventions are more effective for primary and secondary students than other groups, and further, the effect of prejudice reduction interventions endures over time.

Keywords: prejudice, meta-analysis, field experiments, systematic review, effectiveness

Introduction

Fractured intergroup relations and prejudicial attitudes based on ethnic, religious, sexuality and gender-identity differences are a regular feature of social life across many countries (Edelman, Luca, & Svirsky, 2017; Jetten & Esses, 2018; Pedersen & Hartley, 2012; Quillian, Pager, Hexel, & Midtbøen, 2017; Ramsay & Pang, 2017). In the social and behavioural sciences, a significant body of scholarship is concerned with prejudice reduction. However, a 2009 review of this research found that most studies relied on non-experimental methods and laboratory experiments, rather than on field experiments (Paluck & Green, 2009). While these approaches provide valuable information about the types of interventions available, and their effectiveness in abstract laboratory settings, field experiments provide the clearest evidence on what works in real-world settings (Green & Gerber, 2012; Paluck & Green, 2009).

Field experiments can potentially demonstrate the causal mechanisms in a real world setting where broader political and economic influences are at play. Interventions tested in artificial settings may not necessarily yield the same results in the real world where the context is vastly different (Levitt & List, 2007). For the purposes of designing a prejudice reduction intervention, with the goal of real-world application, testing in field environments is critical. Without evidence of interventions that reduce prejudice in real world settings, academics and practitioners are faced with a difficult task of trying to identify the causal mechanisms that work outside of highly controlled and often contrived laboratory settings.

One of the most comprehensive reviews of prejudice reduction interventions was published over a decade ago. Paluck and Green (2009), reviewed 985 published and unpublished prejudice reduction studies. Their review identified a lack of field experiments, leading them to conclude that little was known about which interventions were effective outside of laboratory

settings. Paluck and Green's (2009) review did not extend into meta-analysis, thus the average effect sizes of various interventions and the moderating factors were not identified.

Previous reviews that did contain meta-analysis have been limited in scope. They have either focused on one particular approach such as contact (Lemmer & Wagner, 2015; McIntyre, Paolini, & Hewstone, 2016; Paluck, Green, & Green, 2019; Pettigrew & Tropp, 2006; Zhou, Page-Gould, Aron, Moyer, & Hewstone, 2019), considered particular types of field settings (such as schools only (Ülger, Dette-Hagenmeyer, Reichle, & Gaertner, 2018)), included laboratory studies in addition to field (Lee, Ata, & Brannick, 2014; McIntyre et al., 2016; Pettigrew & Tropp, 2006) or limited their enquiry to specific forms of target prejudice, for example racism (Lee et al., 2014; Lemmer & Wagner, 2015). Certainly, these meta-analyses lend support to the effectiveness of prejudice reduction interventions, yet their scope does not address whether prejudice reduction interventions are in fact effective across a broad range of intervention approaches, field settings, participant groups and prejudice targets.

In addition to understanding the mechanisms of prejudice reduction, it is also important to identify what interventions are likely to have a sustained impact over time. However, there is a lack of review-level evidence about the long-term impacts of prejudice reduction interventions in the field. Only one meta-analysis addressed the issue of persistence, albeit only for contact interventions. It found that there was evidence that contact interventions have sustained impact over time (Lemmer & Wagner, 2015). However, while it focused on field settings, it did not only include randomised experimental approaches.

The purpose of this systematic review and meta-analysis is to extend the work of Paluck and Green (2009) by identifying the characteristics of successful prejudice reduction

interventions in real world settings, across all types of intervention approaches, participants, and prejudice targets.

The key questions guiding this meta-analysis are:

Q1: Do prejudice reduction interventions have causal effect in the field?

Q2: Are some approaches more effective than others?

Q3: Does effectiveness vary across a) participant groups and b) prejudice types?

Q4: What is the persistence of any prejudice reduction effect?

This review covers the period since the Paluck and Green review in 2009. Using a broader scope, this review provides the most comprehensive meta-analysis of prejudice reduction field experiments to date.

Method

A systematic review is a well-defined and transparent approach to synthesising the findings of a body of literature. In comparison to narrative reviews, systematic reviews have the benefit of being less susceptible to selection bias (Uman, 2011). Furthermore, transparent inclusion and exclusion criteria help to overcome the potential for bias in the selection of studies.

Search strategy and inclusion criteria

We used a search strategy based on terms used by Paluck and Green (2009), including: prejudice, attitude, racism, experiment, random*, control*, reduc*, intervention and program. The search covered the time period from 1 January 2009 to 1 January 2019. Three databases were examined: Proquest, Web of Science and PsychINFO. In addition, other unpublished works and grey literature were uncovered through searches on Open Science Framework pre-prints and Google Scholar. The Google Scholar search was limited to results that cited Paluck and Green (2009) in order to increase likelihood of relevance. The final search was conducted in April 2019

and yielded 2912 results. Twenty-seven studies remained after screening (Figure 1). The full list of included studies can be found in the Supplementary Materials.

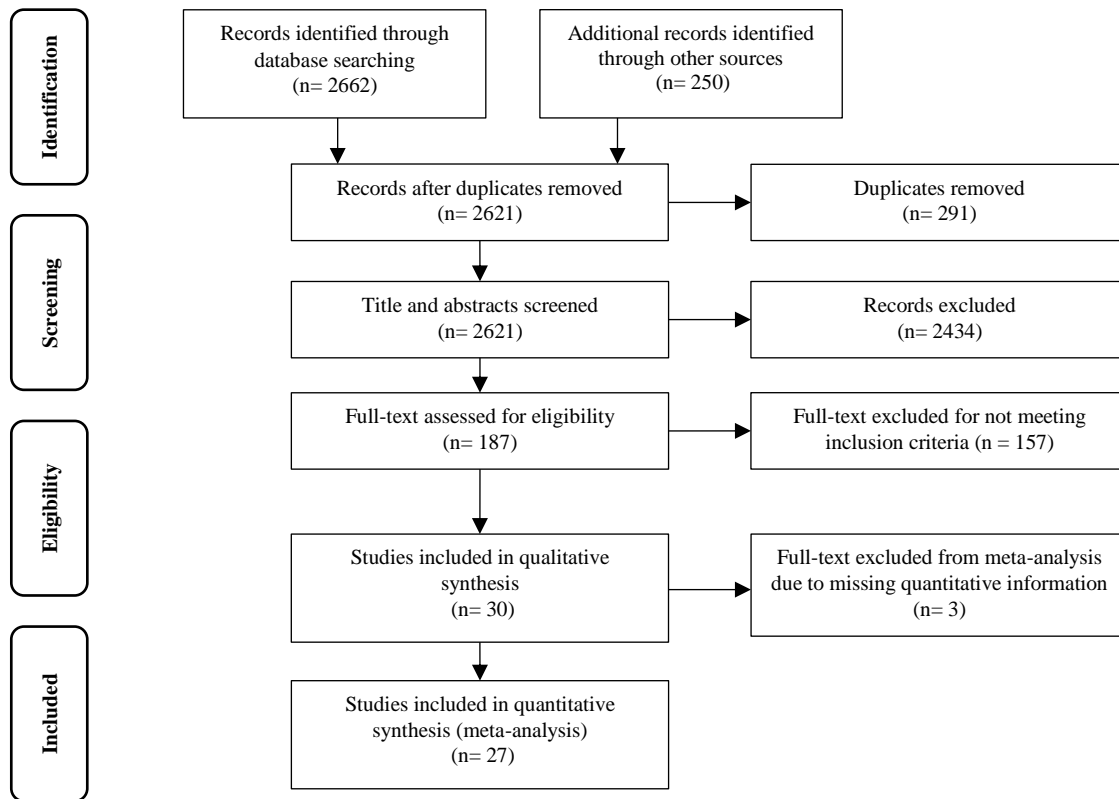


Figure 1. PRISMA flow diagram

The inclusion criteria for the meta-analysis was as follows. First, to be included in the meta-analysis, a paper needed to detail an intervention specifically designed to reduce prejudice. For the purposes of this review, and in line with a significant body of literature, prejudice was defined as negative attitudes towards one group (Allport, 1954; Curseu, Stoop, & Schalk, 2007; Paluck & Green, 2009). Following the approach of Paluck and Green (2009), we excluded studies about sexism because of the distinct theoretical explanations for this form of prejudice (Paluck & Green, 2009). Second, we restricted the timeframe to papers published from 2009 onwards. Third, papers needed to feature field experiments, that is, have a randomised allocation into a control or treatment group, and be set in a field setting. We followed the definition used by

Paluck and Green (2009) that states field experiments are those that are conducted in naturalistic settings in a largely unobtrusive manner that aim to test more context-specific hypotheses.

Studies included in the meta-analysis did not need to have a pre-post design. Fourth, we included all population groups, intervention approaches and control groups. Fifth, papers needed to be in English. A protocol describing our review methodology was preregistered on Open Science Framework ([link](#)). The first author assessed all studies to determine inclusion for the meta-analysis, and discussed any borderline cases with the other authors to determine inclusion.

Coding procedure

We coded information on several study features including: what and when post-intervention measures were collected; participant type (e.g. students, adults); country that the intervention was tested in; sample size; target prejudice type (e.g. racism, homophobia); intervention approach (e.g. contact, perspective taking); effect size and standard error, and details on the prejudice measure used. Coding was completed by the first author.

Our classification of intervention types centered on the underlying theoretical foundations of the intervention in each study. *Contact* interventions were those that relied on contact theory as the mechanism for change in prejudice. This included all forms of contact such as direct / face-to-face, extended contact (e.g. contact is through a form of media such as film or books) and virtual contact (e.g. contact is through non-face-to-face channels, such as instant messaging).¹ *Awareness* interventions were those that sought to reduce prejudice through education. These awareness interventions emphasised the negative consequences of prejudicial views or address the errors associated with stereotypical views or common assumptions. *Categorisation* approaches drew on cognitive recategorisation theories that postulate that the cognitive process

¹ See Lemmer & Wagner (2015) for further details on the types of contact interventions.

of recategorising ingroup and outgroup members as members of an overarching and inclusive superordinate group can reduce prejudice (Dovidio, Gaertner, & Saguy, 2009). Categorisation is often considered along with the dual identity approach whereby both the superordinate category and the original subgroup identity is activated (Dovidio, Gaertner, Pearson, & Riek, 2005).

Perspective taking approaches focused on drawing parallels to a subject's own experiences, creating an experience by which the subject experiences life as an outgroup member and creating empathy with outgroup members as a mechanism for reducing prejudice (Batson et al., 1997; Faulkner, 2018). *Social norm* approaches aimed to challenge norms by creating role models in a community or by highlighting non-prejudiced behaviour through media (Crandall, Eshleman, & O'Brien, 2002). Lastly, *perceived variability* interventions highlighted heterogeneity within an outgroup, that is, that outgroup members have diverse characteristics and are not all 'the same' (Brauer & Er-rafiy, 2011).

During the data extraction process, we clarified additional details related to data extraction. For example, some studies reported on multiple outcomes, included multiple treatment arms or measured outcomes at multiple times. We established the following rules for selecting the main outcome measure to represent each study's overall findings: (a) Where effects were split into majority versus minority groups, only majority group participants were included; where it was not clear which group was the majority, both groups were included; (b) where there were multiple post-tests, we used the interaction of time x group where available or chose the last possible post-test;² (c) where there were multiple intervention arms (i.e. there was more than one treatment group and each treatment group received a different treatment), each intervention

² In one instance, multiple papers were written on the same intervention but tested at different timeframes, so only the paper with latest post-test was included (White & Abu-Rayya, 2012; White, Abu-Rayya, Bliuc, & Faulkner, 2015).

arm was included separately. In addition, where groups received varied dosage of the same treatment, the highest dosage was selected. Similarly, where there was an interaction of treatments (2x2) but it was possible to split effects into two treatments, both treatments were included; (d) Where there were multiple prejudice outcomes, the most general measure, such as one that did not focus on a particular minority was used, or the measure that was closest to “negative attitude towards one group” definition. Where there was a choice between explicit and implicit measures, the explicit measure was used. A preference was given for explicit measure as only one study solely used an implicit measure. The sensitivity to the preference towards explicit measures was tested and will be discussed below.

To facilitate meta-analytic comparison, effect sizes were converted into Cohen’s *d*, which represents the standardised difference between two means (Borenstein & Hedges, 2019).³ Two studies, Belet (2018) and Aronson et al (2016), did not provide enough information to calculate standard effect sizes so were excluded from the meta-analysis. The final sample for this study therefore comprised 27 studies and 34 intervention arms.

Publication bias

We used three different approaches to identify publication bias. First, we plotted standard errors against effect size. Second, we used a contour-enhanced funnel plot (Peters, Sutton, Jones, Abrams, & Rushton, 2008). Finally, because it can be difficult to detect publication bias through visual inspection methods alone (Egger, Smith, Schneider, & Minder, 1997), we used the Egger metabias test. The Egger metabias test is based on a linear regression of normalised effect estimates (estimate divided by its standard error) against precision (reciprocal of the standard error of the estimate) (Egger et al., 1997).

³ Cohen’s *d* were calculated using Practical Meta-Analysis Effect Size Calculator [Online Calculator] (Wilson, n.d.).

Evidential value (p-curve)

We also considered whether there was evidential value in the set of studies – that is, whether these studies indicated a true effect on prejudice or whether the effect may be the result of selective reporting or “p-hacking” (Simonsohn, Nelson, & Simmons, 2014). This was tested through p-curve analysis to assess the distribution of statistically significant p values for a set of independent findings. A right-skewed distribution indicates that there is evidential value in the set of findings, whereas a p-curve skewed to the left suggests intense “p-hacking” or that published results are not representative of the true effect.

Testing for heterogeneity

We expected that effects would be heterogeneous and therefore used a random effects meta-analysis approach. This approach was particularly suitable as it pooled the effects of studies using multiple different prejudice reduction approaches and targeted an outcome, prejudice, which is known to be complex and multidetermined (Cook et al., 1992; Pettigrew & Tropp, 2006). Both Cochran’s Q-test (Cochran, 1954; Hedges & Olkin, 1985) and Higgins’ and Thompson’s I^2 test (Higgins & Thompson, 2002) indicated high degrees of heterogeneity ($Q = 438.78$, $df = 33$, $p < 0.0001$; $I^2 = 92.5\%$), providing additional support for the use of the random effects meta-analysis approach.

Results**Study characteristics**

The meta-analytic dataset contained 34 intervention arms. Study sample sizes ranged from 32 to 1012 ($M = 221$, $SD = 204$, $Mdn = 159$). The dataset included 7522 participants across all treatment arms.

Location.

Studies were spread across the world. Europe accounted for 44 per cent of studies, while Asia and Middle East accounted for 24 per cent (Table 1). Only 15 per cent of field experiments were conducted in the US.

Table 1 Description of included interventions

Subgroup	<i>n</i>	%
<i>Regions</i>		
Europe	15	44
North America	5	15
Asia and Middle East	8	24
Other (inc. Australia, Africa)	6	18
Total	34	100
<i>Participants</i>		
Adults	9	26
Primary	8	24
Secondary	10	29
College	6	18
Adult/students	1	3
Total	34	100
<i>Approach</i>		
Contact	17	50
Awareness	13	38
Categorisation	1	3

Perspective taking	6	18
Social norms	4	12
Perceived variability	1	3
Total	42*	100
<i>Target</i>		
Race/religion/ethnicity	27	77
Obesity	4	12
Homophobia	1	3
Transphobia	1	3
Disability	1	3
Total	34	100

* Some studies used multiple intervention approaches

Study participants.

The vast majority of studies ($n = 24$, 71%) were conducted using school or college student participants. Only 26 per cent of field experiments since 2009 were run entirely with non-school samples.

Target prejudice.

Race-, religion- or ethnicity-based prejudice was investigated in 27 (77%) studies. Obesity was the next most studied form of prejudice (4 studies, 12%). Other types of prejudice considered in the field experiments from the past ten years included homophobia, disability and transphobia.

Intervention approach.

Contact was the most widely-used intervention approach, accounting for 50 per cent of studies. Of the 17 studies that used contact, 10 featured a form of face-to-face contact, while 7 used non-face-to-face contact approaches through imagined contact in an online game, extended contact through narrative stories, or media and virtual contact through online chat. The next most common intervention approach was awareness, which was used in 38 per cent of studies. Other approaches used included categorisation (dual identity, recategorisation), perspective taking, social norms, perceived variability and role models. Eight studies used multiple approaches in one intervention condition (e.g. contact and categorisation simultaneously), hence the total under “Approach” exceeds the total studies.

Preliminary analyses**Publication bias.**

Visual inspection of contour-enhanced funnel plots (Figure 2) revealed some missing studies in the lower left region, that is, small studies with non-significant results. However, the extent of the asymmetry was minimal, suggesting that publication bias was not a material issue for this meta-analysis (Peters et al., 2008).

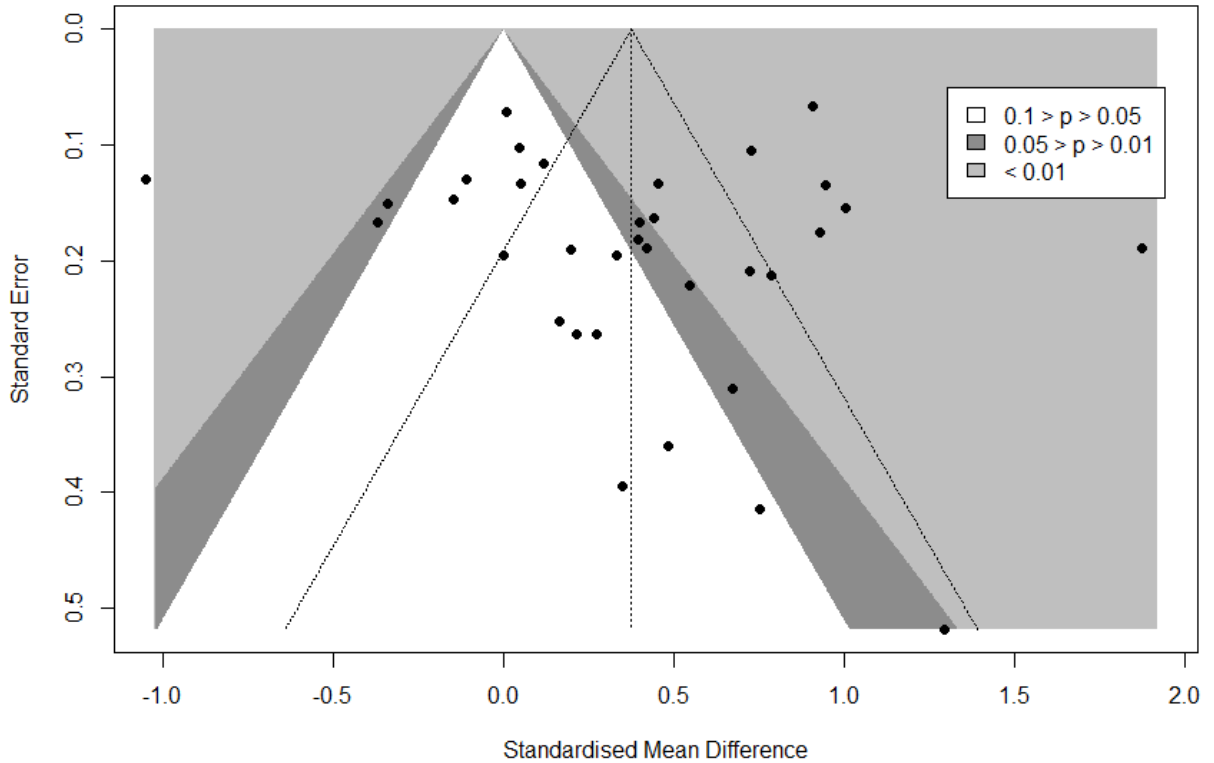


Figure 2. Contour-enhanced Funnel plot

On the plot of standard errors against effect size there appeared to be a positive relationship between smaller standard errors (larger studies) and smaller effect size (i.e. small study effect) suggestive of possible publication bias (Figure 3). However, the formal metabias test supported the null hypothesis of symmetry, indicating no evidence of small study effects ($t = 0.192$, $df = 32$, $p\text{-value} = 0.85$).

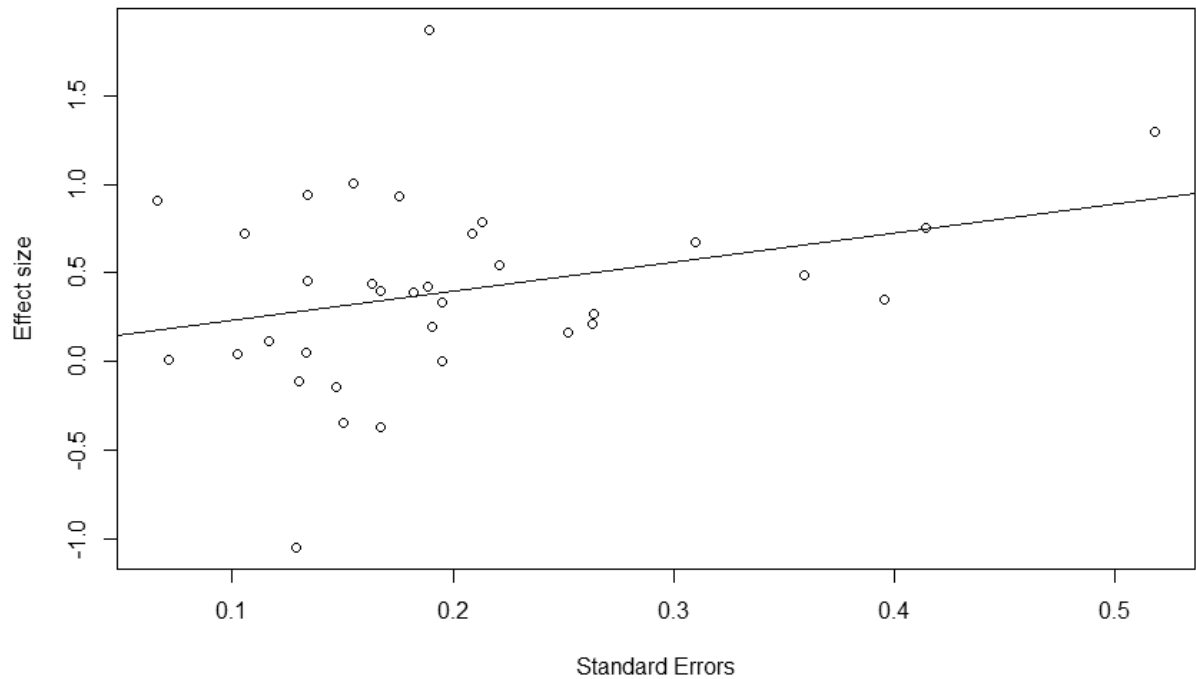
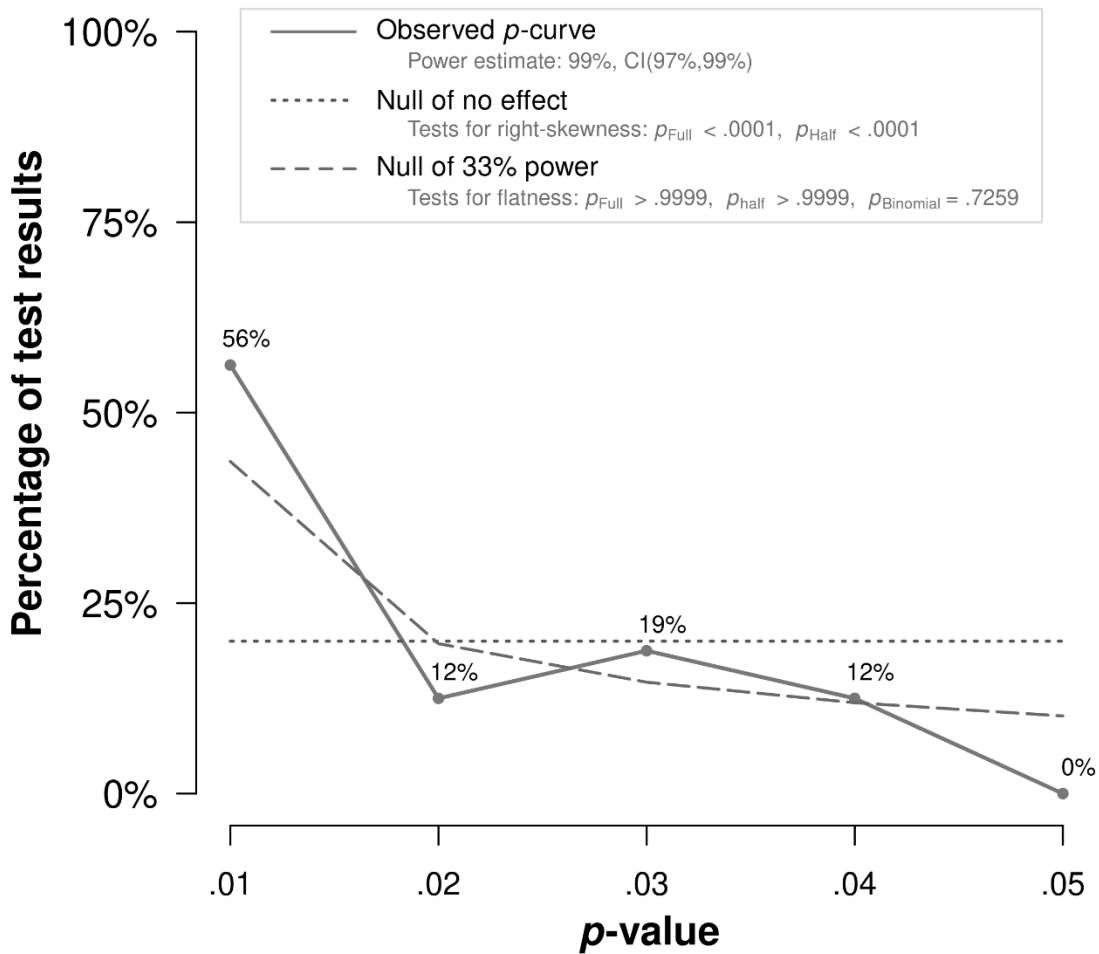


Figure 3. Standard error and effect size

Evidential value (p-curve).

The results of this analysis indicated that there was no evidence of “p-hacking” in the set of studies, with 16 statistically significant results with p-value less than 0.05 and 18 studies with p-value more than 0.05 (Figure 4). Both full and half p-curve tests were right skewed with $p < 0.001$, indicating that there was evidential value (Simonsohn et al., 2014; Simonsohn, Simmons, & Nelson, 2015).



Note: The observed p -curve includes 16 statistically significant ($p < .05$) results, of which 12 are $p < .025$. There were 18 additional results entered but excluded from p -curve because they were $p > .05$.

Figure 4. P-curve (P-curve App 4.06 <http://www.p-curve.com/app4/>, authors' data)

Effect size analyses

Results from the random effects meta-analysis indicated that the average effect of prejudice reduction field experiments was 0.37, with a 95 per cent confidence interval of 0.19 to 0.56 ($p < 0.001$; Figure 5).⁴ On average, prejudice scores were 0.37 standard deviations lower in the intervention group compared with the control. The prediction interval was -0.66 to 1.40. This

⁴ Positive standard mean difference indicates an improvement in attitudes i.e. a reduction in prejudice.

is the range that if a new study was conducted, there would a 95 per cent chance that the standardised mean difference from that study would fall within (Higgins, Thompson, & Spiegelhalter, 2009). In this case the prediction interval was particularly wide because of the high degree of heterogeneity detected in effect size.

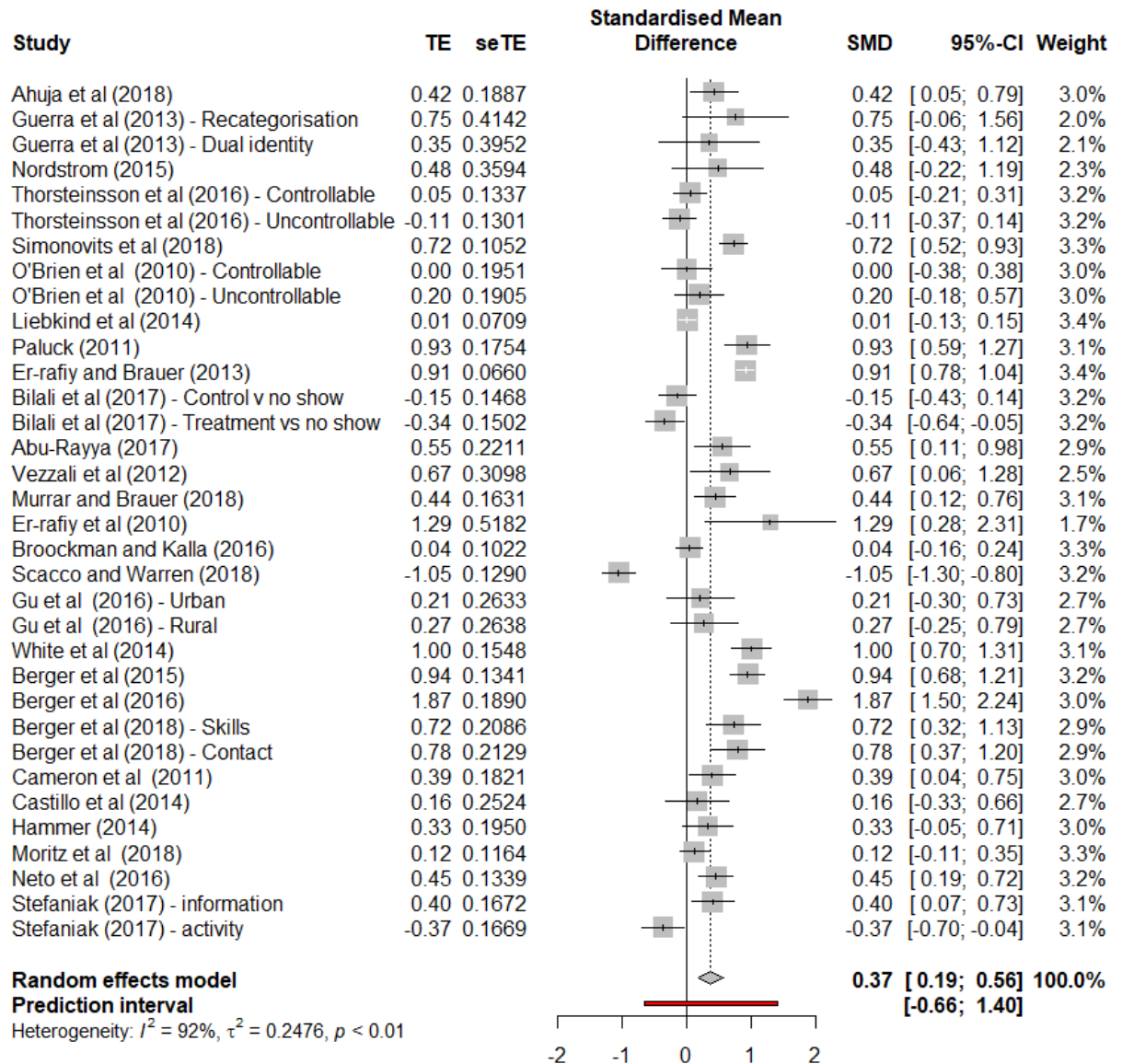


Figure 5. Forest plot of standard errors and effect sizes

Sensitivity test.

We considered whether these results were sensitive to the rules that were used for choosing the relevant prejudice measure, especially our preference for explicit measures. When the meta-analysis was run with a preference for implicit prejudice measures over explicit measures, the average effect size was higher ($SMD = 0.42$, compared to $SMD = 0.37$) and remained statistically significant (95% CI [0.220; 0.615], $p < 0.001$), suggesting that focusing on explicit measures was the more conservative approach to this analysis.

Relative effectiveness of approaches.

To compare effect sizes across intervention approaches we ran subgroup analyses. While there is no strict minimum number of studies required per group, there is greater power to test for subgroup differences the more studies there are in each subgroup. Conclusions based on groupings with very few studies should be interpreted with caution.

When we grouped the studies by approach, the most common approaches were contact and awareness ($n = 10$ for each approach). Nine studies used a combination of multiple approaches (Table 2).

Table 2 Comparison of approaches

Approach	Number of studies (n)	Average effect	Confidence interval
Contact only	10	0.356~	[-0.054; 0.766]
Awareness only	10	0.163	[-0.087; 0.413]
Social norms only	3	0.143	[-1.552; 1.837]

Perceived variability only	1	0.908***	[0.778; 1.037]
Perspective taking only	2	0.464	[-3.081; 4.008]
Contact and Categorisation	3	0.817*	[0.027; 1.607]
Contact, Awareness and Perspective taking	2	1.147	[-8.068; 10.361]
Contact and Perspective taking	1	0.724***	[0.518; 0.930]
Contact and Social norms	1	0.010	[-0.129; 0.148]
Awareness and Perspective taking	1	0.044	[-0.156; 0.244]
Total	34	0.37**	[0.19;0.56]

~p<0.1, *p<0.05, **p<0.01, ***p<0.001

Based on the sub-group analyses, it appears that all approaches and combinations thereof used in field experiments from the past 10 years indicated improved trends in attitudes on average. However, the 95 per cent confidence interval for many subgroups spanned zero, indicating that not all of these effects were statistically significant. Tests for subgroup differences suggested that the differences in effect sizes across approaches was significant ($Q = 121.55$, $df = 9$, $p < 0.001$).

Of the single approach studies, the largest effect size was for the perceived variability approach (0.908) but this was based on only one study. Contact only (0.356), on average, had a larger effect size than awareness only (0.163).

Comparing studies using multiple approaches to studies using single approach suggested that effect trended higher when multiple approaches were used (0.642 vs 0.291). However, this difference was not statistically significant ($Q = 2.15$, $df = 1$, $p\text{-value} = 0.14$).

Relative effectiveness across participant groups.

Subgroup analysis by participants type (primary/elementary school students, secondary/high school students, college, and adults) suggested that prejudice reduction interventions result in an improvement in attitudes for all participant groups (average effects are all positive; Table 3). However, effect sizes appeared to decrease with age, and the impact was more equivocal for older groups (college and adults) where the 95 per cent confidence interval spanned across zero. The difference in average effect size was highly significant ($Q = 16.27$, $df = 4$, $p\text{-value} = 0.003$). Overall these results suggested that prejudice reduction interventions are most effective for primary and secondary students.

Table 3 Subgroup analysis by participants

Approach	Number of studies (<i>n</i>)	Average effect	Confidence interval
Primary	8	0.843**	[0.435; 1.252]
Secondary	10	0.477*	[0.124; 0.830]
College	6	0.073	[-0.571; 0.717]
Adults	9	0.094	[-0.148; 0.336]
Total	34	0.37**	[0.19;0.56]

* $p < 0.05$, ** $p < 0.01$

Relative effectiveness across prejudice types.

We also considered whether the effects of prejudice reduction interventions varied across target prejudice types. The largest effect size was for interventions tackling racism / religious prejudice (0.442) and subgroup analysis suggested that the difference in effect size across target prejudice types was significant ($Q = 13.38$, $df = 3$, $p\text{-value} = 0.004$; Table 4). However, the number of studies was relatively low for the other categories and these results should be interpreted with caution. In fact, when the obesity studies were dropped from this subgroup analysis, the difference across groups was no longer significant ($Q = 1.29$, $df = 2$, $p\text{-value} = 0.52$).

Table 4 Subgroup analysis by target prejudice type

Approach	Number of studies (<i>n</i>)	Average effect	Confidence interval
Race/religion /ethnicity	27	0.442***	[0.212; 0.672]
Obesity	4	0.015	[-0.183; 0.2127]
Sexuality / gender identity	2	0.197	[-2.158; 2.552]
Disability	1	0.393*	[0.036; 0.7498]
Total	34	0.37**	[0.19;0.56]

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Persistence of any prejudice reduction effect.

To examine if the impact of these interventions persisted beyond the immediate context of the intervention, we compared average effect across studies grouped by time of post-measurement. The vast majority of interventions were tested immediately following the

experiment ($n = 13$). Many studies were tested within 1 month of the intervention ($n = 10$). The number of studies tested two months or longer after the intervention dropped off notably (Figure 6).

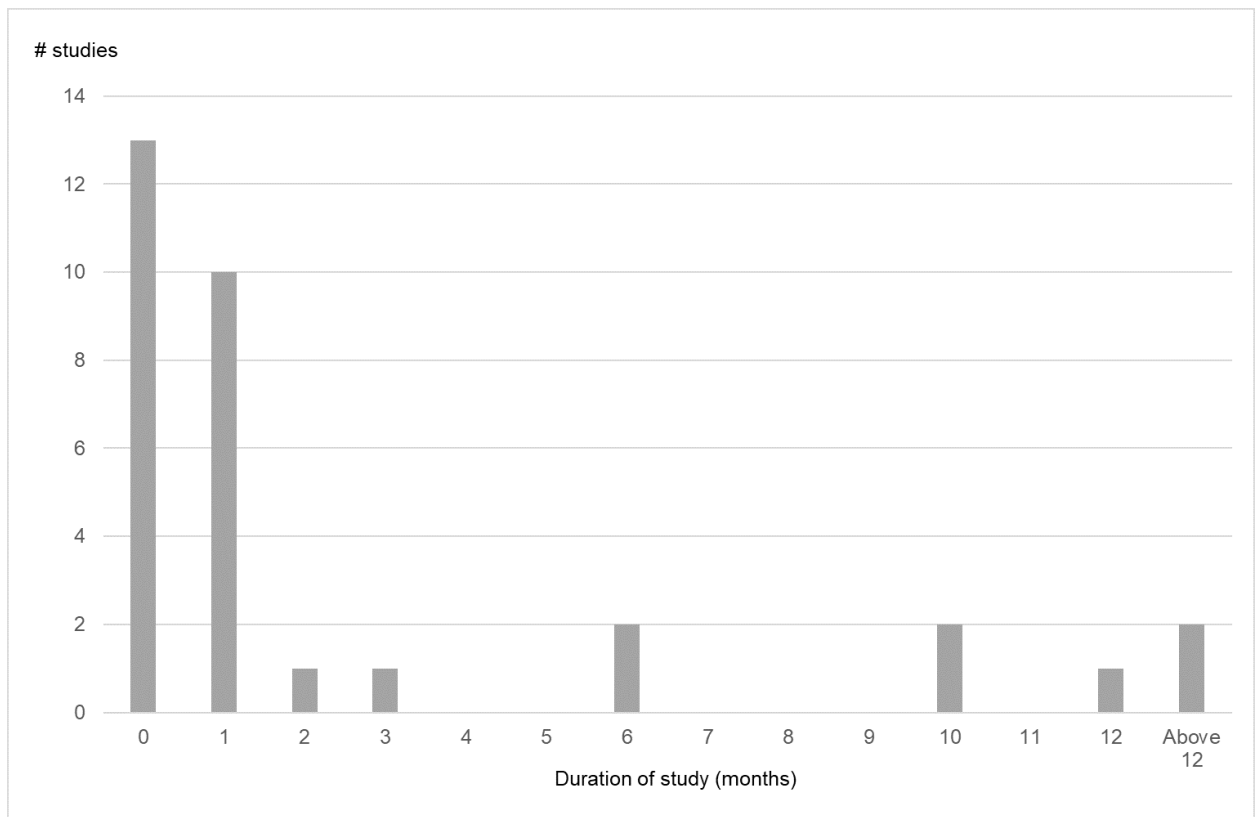


Figure 6. Distribution of duration for post-tests

The average effect size at all three timepoints was positive indicating that there was an improvement in attitudes, regardless of when the post-test was run (Table 5). Although the average effect size trended higher with time, the subgroup analysis showed that the difference in average effect across the three was only marginally significant ($Q = 7.48$, $df = 3$, $p\text{-value} = 0.06$). This provides some evidence that interventions may take time to become maximally effective in reducing prejudice.

Table 5 Subgroup analysis by time of measurement

Approach	Number of studies (n)	Average effect	Confidence interval	Minimum / maximum duration in group (weeks)	Mean (weeks)	Median (weeks)
A – immediately after	13	0.159	[-0.064; 0.381]	0 / 0	0	0
B – up to two months	11	0.480**	[0.233; 0.728]	1 / 6	3.8	0.5
C – over two months	8	0.674*	[0.184; 1.163]	12 / 104	30	24
Total	34	0.37**	[0.19;0.56]	--	--	--

*p<0.05, **p<0.01

Discussion

This systematic review and meta-analysis contribute to knowledge by identifying the average effects of prejudice reduction interventions in real world settings, across all types of interventions, participants, and prejudice targets. To our knowledge, no other review has focused on field experiment evidence across this broad range of sub-categories. The results of this review suggest that prejudice reduction interventions are effective, on average, at improving attitudes towards minority groups in real world settings. We found that less commonly tested approaches, such as perceived variability and perspective taking, had larger effects on average than commonly tested approaches like contact. Interventions also appeared more effective when involving primary or secondary school students. There was strong indication that the impact of interventions tended to be enduring. Additionally, we found some tentative indication that the effects of prejudice reduction intervention may become stronger over time. These findings are discussed in more detail below.

The relative effectiveness of intervention approaches

Subgroup analysis of the various intervention approaches suggested that some approaches were likely to be more effective than others. Between the two most commonly tested approaches, contact and awareness, the meta-analysis results point to stronger effects from contact-based interventions. This difference is not surprising as there were different theories and mechanisms at play in each approach (Brewer, 1999; Hässler et al., 2020; Tajfel, 1969). Contact evokes a number of processes, such as reduced anxiety and increased empathy, in reducing prejudice (Pettigrew, Tropp, Wagner, & Christ, 2011). It could be that contact interventions influence a broader range of mediating causal processes than awareness interventions, which are focused solely on cognition.

Beyond the commonly tested approaches of contact and awareness, this review has identified two promising avenues for theoretical innovation that need to be the focus of further research – perceived variability and perspective taking. Perceived variability is considered to be relatively cheap and simple approach to prejudice reduction in the real world as there are few practical pre-conditions to implementation (Er-rafiy & Brauer, 2012). Further, like contact, perceived variability interventions may also trigger multiple mechanisms associated with prejudice – cognitive, affective and behavioural – potentially resulting in greater impact (Er-rafiy & Brauer, 2013). The results of this review show that the only field trial using a perceived-variability approach had a larger effect size than the average effect of contact or awareness interventions. Similarly, perspective taking – an approach that is particularly effective in increasing empathy, which in turn reduces prejudice (Batson et al., 1997; Faulkner, 2018) – also has a relatively large effect size according to the meta-analysis. Given that research has found that papers reporting novel effects tend to overestimate the size of those effects (Klein et al.,

2018), more research is needed to identify the average effect of perceived variability and perspective taking interventions. As such, while these interventions appear promising, more research is needed to investigate their effects in real-world settings. Consistent with research on related topics, we found that effect sizes trended marginally higher in interventions that used multiple approaches simultaneously than in interventions that relied on a single approach (Bezrukova, Spell, Perry, & Jehn, 2016; Ülger et al., 2018). It could be using multiple approaches makes interventions better able to tackle the complexity of prejudice in the real world and minimise the impact of the “single factor fallacy” (Pettigrew & Hewstone, 2017). Although it is important to test individual approaches to identify their unique effects on prejudice, these results indicate that interventions combining multiple theoretical approaches may have a larger impact on prejudice, and suggest that research that tests combinations of intervention types is needed.

Schools appear to be more effective settings

The types of participants featured in field experiments remain largely unchanged from the 2009 review by Paluck and Green. In the last ten years, field experiments on adults in the community have not gained momentum. Instead, the vast majority of prejudice reduction experiments are concentrated in educational settings. Within the school context, prejudice reduction interventions appeared more effective for primary and secondary students compared to for college and adults. This was broadly consistent with the findings of other meta-analyses on specific interventions types (Pettigrew & Tropp, 2006).

There are at least three plausible reasons for this finding. First, there are higher degrees of control of the environmental conditions in primary and secondary school settings and this may lead to greater intervention effectiveness. This is supported by prior research that finds

environmental conditions enhance the impact of contact on prejudice (Pettigrew & Tropp, 2006). First, the original form of contact theory specified four conditions that needed to be met for contact to be effective (Allport, 1954): equal status; common goals; institutional support; and intergroup cooperation. Educational environments are more likely to meet these requirements as students tend to be in fairly egalitarian environments, the curriculum may involve working together in groups, and schools tend to maintain authority, although this cannot be assumed in all instances (Ülger et al., 2018). Second, younger participants may be more receptive to interventions. Developmental psychology research suggests that prejudice is more entrenched among adults than for younger cohorts (Rutland & Killen, 2015; Sears, 1986). Third, it could also be that as much research has been focused on students in school settings, at the expense of research on adults, less is known about what makes an effective community-based adult targeted intervention. While school settings technically meet the definition of a field setting, it is a more controlled environment compared to general community settings, thus, there is greater likelihood of intervention robustness in schools. The smaller effect size for adults could reflect nascent research on how to effectively shift prejudice in less structured communities. Future work should be directed towards a richer understanding of what works in general community settings with adult participants.

Interventions are most effective for prejudice based on race, ethnicity, and religion

The average effect size was largest for studies tackling racism, ethnicity, or religious-based prejudice. This finding is in contrast to previous meta-analyses. Previous meta-analyses, focusing only on contact interventions, either found that effect size was larger for other types of prejudice (Paluck et al., 2019) or that effect size was not significantly different across categories (Pettigrew & Tropp, 2006). However, this could be a consequence of the vast majority of studies

being on racism or religion. This conclusion was strengthened by the result that when the obesity studies were dropped as outliers, that the difference across the different prejudice types was not significant.

Sustained impact of prejudice reduction interventions

We looked at field experiments because we wanted to examine what works in the real world to inform practical efforts to reduce prejudice. Thus, it was critical to know whether effects endured over time. Promisingly, results indicate that the impact of prejudice reduction interventions endured overtime. This finding is consistent with Lemmer and Wagner (2015), who found that contact effects persisted over time. However, caution should be exercised in interpreting this conclusion as the analysis showed that few studies were tested two months or more after the intervention, presenting a significant gap in the literature on what happens after this period. The gap was also recently noted in the context of prejudice reduction interventions in school contexts (Ülger et al., 2018).

Not only did the effects of prejudice reduction interventions persist over time, we also found a weak, marginally significant, indication that the impact of these interventions became larger over time. This result adds richness to what we understand about persistence from single studies. Looking at single studies suggested that the persistence of prejudice reduction interventions was mixed. Multiple studies found a significant but smaller impact on prejudice over time – ranging from 4-6 weeks to 15 months after the intervention (Berger, Benatov, Abu-Raiya, & Tadmor, 2016; Murrar & Brauer, 2018; White, Abu-Rayya, & Weitzel, 2014). While other studies found a significantly stronger effect over time periods of three months to two years after the intervention (Berger, Brenick, Lawrence, Coco, & Abu-Raiya, 2018; Neto, Da

Conceição Pinto, & Mullet, 2016) or a non-significantly stronger effect over time (Broockman & Kalla, 2016). These mixed results could reflect a number of different mechanisms at play.

The marginally significantly increase in the effects on prejudice over time could be due to changes in how individuals interpret new information about the relevant outgroup.

Research has indicated that prejudice against an outgroup increases the likelihood of relying on misinformation about that outgroup (Ecker, Lewandowsky, Fenton, & Martin, 2014). As such, reducing prejudice could lead to further reductions in prejudice over time because it changes subsequent interpretation of information about the relevant outgroup. Further research is needed to investigate this possibility.

More generally, the strengthening of the impact of prejudice reduction intervention over time warrants further investigation in future studies. A larger sample of studies with outcomes measures at multiple points is necessary to give a clearer picture of whether the effect size of prejudice reduction interventions changes over time and the relevant mechanism at play.

Scalability of prejudice reduction interventions

Given the pernicious problem of prejudice in the community, we note that discussions pertaining to the scalability of interventions, or the capacity of programs and interventions to increase in reach and impact, were almost universally absent in published literature. Of the studies that we considered in this meta-analysis, there was no explicit and detailed consideration of how scalable effective interventions may be to larger populations. For interventions to have a sustained impact in the real world, implementation science research indicates that it is not enough to just consider effectiveness, it is also necessary to consider scalability (Milat, King, Bauman, & Redman, 2013). Consideration of scalability is also essential for practitioners who may be interested in replicating the intervention in other contexts. Future research is needed to

identify the characteristics that prejudice reduction interventions need to be scalable, and consideration should also be given as to whether particular prejudice reduction interventions meet those criteria.

Limitations

The primary limitation of this review is that some of the subgroup analyses relied on a small number of studies. For example, there were a relatively small number of studies that: tested interventions other than contact and awareness; used adult participants; or, focused on prejudice based on factors other than race, ethnicity, or religion. Although we included all available studies that met inclusion criteria, there remains a need for subsequent research on these topics so that average effects can be more accurately estimated.

Conclusion

This systematic review and meta-analysis show that research on prejudice reduction has matured substantially since Paluck and Green's (2009) review identified a dearth of field experiments on the topic. Our meta-analysis of 27 field experiments published since 2009 indicates that prejudice reduction interventions are effective on average in reducing prejudice in real world settings. Moreover, their effects tend to persist over time. However, our review identifies that the effectiveness of prejudice reduction interventions varies across intervention types and participant types. Prejudice reduction interventions have larger effect size when using contact approaches over awareness approaches, and a larger effect for primary and secondary students versus college students and adults.

Notwithstanding these advances, there is still need for more research using field experiments if researchers and practitioners want to successfully tackle prejudice in the real world. It is necessary to continue to build evidence in a way that helps understand the

effectiveness of the suite of theories and methods available. Future work should consider the relative effectiveness of different approaches and theories especially those approaches that have had promising initial findings but relatively little research, such as perceived variability. Similar focus should be placed on the relatively unexplored participant group of adults in community settings.

References

- Allport, G. W. (1954). *The nature of prejudice*. Cambridge MA: Addison-Wesley.
- Aronson, K. M., Stefanile, C., Matera, C., Nerini, A., Grisolaghi, J., Romani, G., ... Brown, R. (2016). Telling tales in school: Extended contact interventions in the classroom. *Journal of Applied Social Psychology, 46*(4), 229–241. <https://doi.org/10.1111/jasp.12358>
- Batson, C. D., Polycarpou, M. P., Harmon-Jones, E., Imhoff, H. J., Mitchener, E. C., Bednar, L. L., ... Highberger, L. (1997). Empathy and attitudes: Can feeling for a member of a stigmatized group improve feelings toward the group? *Journal of Personality and Social Psychology, 72*(1), 105–118. <https://doi.org/10.1037/0022-3514.72.1.105>
- Belet, M. (2018). Reducing interethnic bias through real-life and literary encounters: The interplay between face-to-face and vicarious contact in high school classrooms. *International Journal of Intercultural Relations, 63*, 53–67. <https://doi.org/10.1016/j.ijintrel.2018.01.003>
- Berger, R., Benatov, J., Abu-Raiya, H., & Tadmor, C. T. (2016). Reducing prejudice and promoting positive intergroup attitudes among elementary-school children in the context of the Israeli–Palestinian conflict. *Journal of School Psychology, 57*, 53–72. <https://doi.org/10.1016/j.jsp.2016.04.003>
- Berger, R., Brenick, A., Lawrence, S. E., Coco, L., & Abu-Raiya, H. (2018). Comparing the effectiveness and durability of contact- and skills-based prejudice reduction approaches. *Journal of Applied Developmental Psychology, 59*, 46–53. <https://doi.org/10.1016/j.appdev.2018.04.002>
- 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),

1227–1274. <https://doi.org/10.1037/bul0000067>

Borenstein, M., & Hedges, L. V. (2019). Effect sizes for meta-analysis. In H. Cooper, L. V.

Hedges, & J. C. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-Analysis*

(Vol. 38, pp. 207–242). <https://doi.org/10.2105/ajph.38.9.1289-b>

Brauer, M., & Er-rafiy, A. (2011). Increasing perceived variability reduces prejudice and

discrimination. *Journal of Experimental Psychology*, *47*, 871–881.

<https://doi.org/10.1016/j.jesp.2011.03.003>

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of*

Social Issues, *55*(3), 429–444. <https://doi.org/10.1111/0022-4537.00126>

Broockman, D., & Kalla, J. (2016). Durably reducing transphobia: A field experiment on door-

to-door canvassing. *Science*, *352*(6282), 220–224. <https://doi.org/10.1126/science.aad9713>

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*,

10, 101–129. <https://doi.org/10.2307/3001666>

Cook, T. D., Cooper, H., Cordray, D. S., Hartman, H., Hedges, L. V., & Light, R. J. (1992).

Some generic issues and problems for meta-analysis. In T. D. Cook, D. S. Cooper, H.

Cordray, H. Hartman, L. V. Hedges, & R. J. Light (Eds.), *Meta-analysis for explanation: A*

casebook (pp. 283–320). New York: Russell Sage Foundation.

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and

suppression of prejudice: The struggle for internalization. *Journal of Personality and Social*

Psychology, *82*(3), 359–378. <https://doi.org/10.1037/0022-3514.82.3.359>

Curseu, P. L., Stoop, R., & Schalk, R. (2007). Prejudice toward immigrant workers among Dutch

employees: integrated threat theory revisited. *European Journal of Social Psychology*,

37(1), 125–140. <https://doi.org/10.1002/ejsp.331>

- Dovidio, J. F., Gaertner, S. L., Pearson, A. R., & Riek, B. M. (2005). Social identities and social context: Attitudes and personal well-being. In S. R. Thye & E. J. Lawler (Eds.), *Advances in group processes: Social identification processes in groups* (pp. 231–260). Oxford: Elsevier.
- Dovidio, J. F., Gaertner, S. L., & Saguy, T. (2009). Commonality and the complexity of “we” social attitudes and social change. *Personality and Social Psychology Review*, *13*(1), 3–20.
<https://doi.org/10.1177/1088868308326751>
- Ecker, U. K. H., Lewandowsky, S., Fenton, O., & Martin, K. (2014). Do people keep believing because they want to? Preexisting attitudes and the continued influence of misinformation. *Memory & Cognition*, *42*(2), 292–304. <https://doi.org/10.3758/s13421-013-0358-x>
- Edelman, B., Luca, M., & Svirsky, D. (2017). Racial discrimination in the sharing economy: Evidence from a field experiment. *American Economic Journal: Applied Economics*, *9*(2), 1–22. <https://doi.org/10.1257/app.20160213>
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*, 629–634.
<https://doi.org/10.1136/bmj.315.7109.629>
- Er-rafiy, A., & Brauer, M. (2012). Increasing perceived variability reduces prejudice and discrimination : Theory and application. *Social and Personality Psychology Compass*, *12*(6), 920–935. <https://doi.org/10.1111/spc3.12000>
- Er-rafiy, A., & Brauer, M. (2013). Modifying perceived variability: four laboratory and field experiments show the effectiveness of a ready-to-be-used prejudice intervention. *Journal of Applied Social Psychology*, *43*(4), 840–853.
<https://doi.org/http://dx.doi.org/10.1111/jasp.12010>
- Faulkner, N. (2018). “Put yourself in their shoes”: Testing empathy’s ability to motivate

cosmopolitan behavior. *Political Psychology*, 39(1), 217–228.

<https://doi.org/10.1111/pops.12411>

Green, D. P., & Gerber, A. S. (2012). *Field Experiments*. New York: W. W. Norton.

Hässler, T., Ullrich, J., Bernardino, M., Shnabel, N., Laar, C. Van, Valdenegro, D., ... Ugarte, L.

M. (2020). A large-scale test of the link between intergroup contact and support for social change. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-019-0815-z>

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, Florida: Academic Press.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21, 1539–1558. <https://doi.org/10.1002/sim.1186>

Higgins, J. P. T., Thompson, S. G., & Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(1), 137–159. <https://doi.org/10.1111/j.1467-985X.2008.00552.x>

Jetten, J., & Esses, V. M. (2018). The reception of immigrants and refugees in Western countries: The challenges of our time. *Journal of Social Issues*, 74(4), 662–673. <https://doi.org/10.1111/josi.12292>

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., ... Nosek, B. A. (2018). Many labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, 1(4), 443–490. <https://doi.org/10.1177/2515245918810225>

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

- 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. <https://doi.org/10.1002/ejsp.2079>
- Levitt, S. D., & List, J. A. (2007). Viewpoint: On the generalizability of lab behaviour to the field. *Canadian Journal of Economics, 40*(2), 347–370. <https://doi.org/10.1111/j.1365-2966.2007.00412.x>
- McIntyre, K., Paolini, S., & Hewstone, M. (2016). Changing people's views of outgroups through individual-to-group generalisation: meta-analytic reviews and theoretical considerations. *European Review of Social Psychology, 27*(1), 63–115. <https://doi.org/10.1080/10463283.2016.1201893>
- Milat, A. J., King, L., Bauman, A. E., & Redman, S. (2013). The concept of scalability: Increasing the scale and potential adoption of health promotion interventions into policy and practice. *Health Promotion International, 28*(3), 285–298. <https://doi.org/10.1093/heapro/dar097>
- Murrar, S., & Brauer, M. (2018). Entertainment-education effectively reduces prejudice. *Group Processes and Intergroup Relations, 21*(7), 1053–1077. <https://doi.org/10.1177/1368430216682350>
- Neto, F., Da Conceição 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. <https://doi.org/10.1177/0305735614568882>
- Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual Review of Psychology, 60*(1), 339–367. <https://doi.org/10.1146/annurev.psych.60.110707.163607>

- Paluck, E. L., Green, S. A., & Green, D. P. (2019). The contact hypothesis re-evaluated. *Behavioural Public Policy*, 3(2), 129–158. <https://doi.org/10.1017/bpp.2018.25>
- Pedersen, A., & Hartley, L. K. (2012). Prejudice against Muslim Australians: The role of values, gender and consensus. *Journal of Community & Applied Social Psychology*, 22, 239–255. <https://doi.org/10.1002/casp>
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, 61(10), 991–996. <https://doi.org/10.1016/j.jclinepi.2007.11.010>
- Pettigrew, T. F., & Hewstone, M. (2017). The single factor fallacy: Implications of missing critical variables from an analysis of intergroup contact theory. *Social Issues and Policy Review*, 11(1), 8–37. <https://doi.org/10.1111/sipr.12026>
- Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90(5), 751–783. <https://doi.org/10.1037/0022-3514.90.5.751>
- Pettigrew, T. F., Tropp, L. R., Wagner, U., & Christ, O. (2011). Recent advances in intergroup contact theory. *International Journal of Intercultural Relations*, 35(3), 271–280. <https://doi.org/10.1016/j.ijintrel.2011.03.001>
- Quillian, L., Pager, D., Hexel, O., & Midtbøen, A. H. (2017). Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. *Proceedings of the National Academy of Sciences of the United States of America*, 114(41), 10870–10875. <https://doi.org/10.1073/pnas.1706255114>
- Ramsay, J. E., & Pang, J. S. (2017). Anti-immigrant prejudice in rising East Asia: A stereotype

content and integrated threat analysis. *Political Psychology*, 38(2), 227–244.

<https://doi.org/10.1111/pops.12312>

Rutland, A., & Killen, M. (2015). A developmental science approach to reducing prejudice and social exclusion: Intergroup processes, social-cognitive development, and moral reasoning.

Social Issues and Policy Review, 9(1), 121–154. <https://doi.org/10.1111/sipr.12012>

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*,

51(3), 515–530. <https://doi.org/10.1037/0022-3514.51.3.515>

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer.

Journal of Experimental Psychology: General, 143(2), 534–547.

<https://doi.org/10.1037/a0033242>

Simonsohn, U., Simmons, J., & Nelson, L. (2015). Better p-curves: Making p-curve analysis

more robust to errors, fraud, and ambitious p-hacking. *Journal of Experimental Psychology:*

General, 144(6), 1146–1152.

Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Social Issues*, 25(4), 79–97.

<https://doi.org/10.1111/j.1540-4560.1969.tb00620.x>

Ülger, Z., Dette-Hagenmeyer, D. E., Reichle, B., & Gaertner, S. L. (2018). Improving outgroup

attitudes in schools: A meta-analytic review. *Journal of School Psychology*, 67, 88–103.

<https://doi.org/10.1016/j.jsp.2017.10.002>

Uman, L. S. (2011). Systematic reviews and meta-analyses. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 20(1), 57–59.

<https://doi.org/10.1016/j.revmed.2014.05.011>

White, F. A., & Abu-Rayya, H. M. (2012). A dual identity-electronic contact (DIEC) experiment

promoting short- and long-term intergroup harmony. *Journal of Experimental Social Psychology*, 48(3), 597–608. <https://doi.org/10.1016/j.jesp.2012.01.007>

White, F. A., Abu-Rayya, H. M., Bliuc, A. M., & Faulkner, N. (2015). Emotion expression and intergroup bias reduction between Muslims and Christians: Long-term Internet contact. *Computers in Human Behavior*, 53, 435–442. <https://doi.org/10.1016/j.chb.2015.04.074>

White, F. A., Abu-Rayya, H. M., & Weitzel, C. (2014). Achieving twelve-months of intergroup bias reduction: The dual identity-electronic contact (DIEC) experiment. *International Journal of Intercultural Relations*, 38(1), 158–163. <https://doi.org/10.1016/j.ijintrel.2013.08.002>

Wilson, D. B. (n.d.). Practical meta-analysis effect size calculator [Online Calculator]. Retrieved November 30, 2019, from <https://campbellcollaboration.org/research-resources/effect-size-calculator.html>

Zhou, S., Page-Gould, E., Aron, A., Moyer, A., & Hewstone, M. (2019). The extended contact hypothesis: A meta-analysis on 20 years of research. *Personality and Social Psychology Review*, 23(2), 132–160. <https://doi.org/10.1177/1088868318762647>