WORKING PAPER SERIES

Contact Interventions: A Meta-Analysis

Gwen-Jiro Clochard

Contact Interventions: A Meta-Analysis[*]

Gwen-Jirō Clochard

July 12, 2022

Abstract

For decades, intergroup contact has been viewed as one of the main tools to reduce prejudice and improve intergroup relations. This paper reviews the experimental literature on the contact hypothesis. Based on an analysis of 62 measures from 37 papers, the conclusions are threefold. First, contact interventions are, on average, effective at reducing prejudice. Second, there exists a very large heterogeneity in the type of interventions labelled as contact. Third, characteristics of the experimental context, rather than the intervention itself, seem to matter for the efficacy of contact. Implications for the future of the contact literature are discussed.

Keywords: contact hypothesis, meta-analysis, prejudice reduction, field experiments

JEL Codes: C93, C12, C83

1 Introduction

Because of its consequences on social inequality (Durante and Fiske, 2017), xenophobia (Kumar et al., 2011) or reduced economic output (Hjort, 2014), solutions to reduce prejudice have been studied in psychology, sociology and economics for decades. Among candidate solutions, *contact* interventions have received the most attention (Bertrand and Duflo, 2017).

The so-called contact hypothesis was first coined by Allport (1954), who posited that prejudice "may be reduced by equal status contact between majority and minority groups in the pursuit of common goals. The effect is greatly enhanced if this contact is sanctioned by institutional supports (i.e., by law, custom, or local atmosphere), and provided it is of a sort that leads to the perception of common interests and common humanity between members of the two groups." (p.281). Since Allport, interventions promoting contact between groups have been seen as one of the main tools to reduce prejudice (Bertrand and Duflo, 2017; Paluck et al., 2021).

The literature devoted to the investigation of the contact hypothesis developed dramatically after Allport's book, with the seminal meta-analysis by Pettigrew and Tropp (2006) identifying no less than 515 studies, covering more than 250,000 people from 38 countries, over a period ranging from the 1940s to the year 2000. The overwhelming conclusion was that contact is effective at reducing prejudice, noting that "Results from the meta-analysis conclusively show that intergroup contact can promote reductions in intergroup prejudice" (p.751).

However, only a small fraction analyzed in the meta-analysis employed an experimental design, with a clear definition of a treated and control group. After the removal of non-experimental protocols - and the addition of more recent papers investigating contact, Paluck et al. (2019) are left with 27 studies investigating contact. Of these, the largest share (33%) investigated the effect of contact on racial and ethnic prejudice for university students or young adults (18-25 years old), and only two were conducted in developing countries - Scacco and Warren (2018) in Nigeria and Corno et al. (2019) in South Africa.

In this paper, I update the analysis by Paluck et al. (2019) by adding new papers on the expanding field of contact interventions. I identified 62 measures from 37 papers. I also deepen the analysis by investigating which characteristics of interventions appear to be associated with a larger impact. Allport (1954) identified four necessary conditions of effectiveness of contact interventions - equal status among groups, common coals, positive contact and the support of authorities. Subsequent work (Pettigrew and Tropp, 2006; Lemmer and Wagner, 2015) also identified friendship potential and scriptedness of encounter as potential mediators of the effect of contact. Importantly, no formal test of the relevant hypotheses has, to the best of my knowledge, ever been performed. I therefore estimate the relevance of characteristics of the interventions, as well as variables indicative of the context of the paper (e.g. prejudice on race/ethnicity or caste), for the

efficacy of contact interventions.

My main results are threefold. The first result is confirming that of Paluck et al. (2019) by showing that contact interventions *are* effective at reducing prejudice and improving measures of intergroup cohesion. The typical intervention will have an effect of approximately 0.33 standard deviations. According to Cohen (1969)'s rule of thumb, contact interventions therefore typically have a small to moderate effect on prejudice.[1] There exist, however, a large heterogeneity between studies and measures.

The second result is that there exists a lack of consistency in use of the term "contact". Just like there exists some debate about what falls under the umbrella of intergroup conflict (Lee and Salvatore, 2022), the term contact has been used to describe very different protocols, ranging from short face-to-face discussions with very scripted protocols (Page-Gould et al., 2008; Clochard, 2021; Clochard et al., 2022), sports leagues lasting several months (Mousa, 2020; Lowe, 2021), education programs (Sorensen, 2010; Scacco and Warren, 2018) to interactions between army recruits over boot camps (Carrell et al., 2015; Finseraas et al., 2016). The variety of protocols renders difficult the exercise of understanding the determinants of efficacy of contact interventions.

The third result is on the characteristics of the most effective contact interventions. The analysis consisted of a Lasso and post-lasso OLS estimation, with the effect size as the dependent variable, and all characteristics of the papers as regressors. Results indicate that the largest effects are found for studies investigating prejudice against older people and people with disabilities. The effect is lower when no physical encounter with a person was run, and when the outcome is measured 1 to 30 days after the end of the intervention. Interestingly, no condition identified by Allport (1954), Pettigrew and Tropp (2006) or Lemmer and Wagner (2015) are found to be strong predictors of the effect. Although these conditions are almost never explicitly randomized,[2] these results indicate that these conditions are neither necessary nor sufficient for an effective contact.

After the presentation of the results, I discuss the implications of this meta-analysis for the future on the research on the contact hypothesis.

The remainder of the paper is organized as follows. Section 2 presents the method of selection of the papers in this analysis, describes the variables of interest and presents the methodology of analysis. Section 3 presents the results. Section 4 discusses the implications of the findings and concludes on future for research on the contact hypothesis.

---

[1]Cohen (1969) identifies effects with $d = 0.2$ as small, $d = 0.5$ as moderate, $d = 0.8$ as large.

[2]One significant exception is Lowe (2021) who randomizes the common goal condition.

2 Method

2.1 Paper selection

The aim of this paper is to investigate the effect of contact interventions on intergroup prejudice. In order to be included in the present analysis, studies had to involve 1) an *experimental* induction of 2) in-person contact, with clearly defined, 3) existing groups.

The experimental criterion was one of the main reason for excluding papers from the analysis. It implies that studies involving quasi-experimental variations (Vertier and Viskanic, 2018; Rao, 2019; Steinmayr, 2021) were not included in the analysis. Studies with no random assignment at all (Alesina et al., 2003; Danckert et al., 2017) were also excluded from the analysis.

The second criterion for inclusion was in-person contact. A second group of excluded studies, which is rapidly growing in quantity (in particular since the COVID-19 pandemic), regroups studies involving online encounters, such as Lenz and Mittlaender (2022).[3]

The third criterion was the application of an intervention on real groups, therefore excluding studies involving artificial groups. These typically include many laboratory experiments, as for instance Whitt et al. (2021), which induce conflict between groups formed during the experiment.

The papers were selected from recent meta-analyses (Lemmer and Wagner, 2015; Paluck et al., 2019), and from Google Scholar searches. For Scholar searches, all papers citing the meta-analyses or Allport's book (after 2015) were searched, as well as a word search for *contact* and *prejudice*. After the application of the three criteria of inclusion, I was left with 37 papers, with publications between 1972 and 2022, spanning nearly all continents and covering, in total, more than 17,000 individuals. The full list of papers is presented in Table A.1.

Importantly, for all selected papers, I included the main outcome variables included from the articles. For instance, in Mousa (2020), I used whether participants attended an event with Muslim players, whether they voted for a Muslim to receive an award and whether they trained with a Muslim six months after the experiment. This distinction was made because some papers have outcomes measuring different things: in Clochard (2021), I investigate the effect of contact separately on trust the specific police officers met, but also toward the police in general. Other papers reported measures separate in time, and were also included. In total, this left 62 measures from the 37 papers.

---

[3]For a meta-analysis of contact in online contexts, see Imperato et al. (2021), who find a positive effect of online contact with outgroup members.

2.2 Description of variables

There are two broad sets of variables used in this analysis: variables related to the contexts of the paper, and variables related to the contact intervention itself. Descriptive statistics are presented in the following Section.

Variables on contexts For the variables on papers and their contexts, I define six variables of interest, which I categorize as follows.

Publication year: I split the sample in three categories: before 2000, between 2001 and 2010 and after 2011.

Sample size: I used four categories: [0,50], [51,100], [101,500] and 501+.

Average age: The variable uses the average age of participants provided by the paper. Three main categories were identified, 0-18 years old, 18-25 and 25+. These categories can be broadly thought of as corresponding to "Children to high-school students", "University students or young adults" and "General adult population".

Zone: I divided papers according to geographical areas. This category includes Asia, Middle-East and Northern Africa, North America, Oceania, Sub-Saharan Africa and Western Europe.

Type of prejudice: This variable captures the main dimension of prejudice targeted by the contact intervention in the paper.[4] The variable is divided in ten categories: age, caste, disabilities, gender, immigrants, LGBTQ+, police, prisoners, race / ethnicity and religion.

Type of outcome: This variable defines the type of outcome used in the paper. This variable is divided into three categories. The first category is behavior (or actions), corresponding to observed actions by participants toward the outgroup. This category can range from experimental games (Finseraas et al., 2019; Clochard, 2021) to the number of friends from another group (DeVries et al., 1977) and the number of emails exchanged with outgroup members (Marmaros and Sacerdote, 2006). The second type of outcome is explicit beliefs or attitudes about the other group. This typically involves participants to declare whether they agree with a pre-defined set of statements explicitly about the other group, e.g. "Affirmative action in college admissions should be abolished" (Boisjoly et al., 2006), "Disabled people are often grumpy and moan about everything" (Krahé and Altwasser, 2006). The last outcome category is implicit behavior. The main outcome used in this case is some version of the Implicit Association Test (Greenwald et al., 1998), for which participants have to click as fast as possible to associate positive or negative words between the different groups. I also defined a dummy variable called *Measure for the entire group* to distinguish measures involving the specific individuals met by participants and the entire outgroup.

---

[4]For instance, in Clochard et al. (2022) the contact appears to be more effective for older individuals and people with lower education, but the main prejudice dimension studied is inter-ethnic trust.

Variables on contact intervention   The contact interventions are widely heterogeneous. I identified several characteristics, which were coded as the following.

Type of contact intervention: The variable consists in broad categories of contact interventions. Seven categories are defined: army recruits, classmates (or participation in a similar course), participation in a collaborative task, discussions (from scripted discussions as in Page-Gould et al. (2008) to door-to-door canvassing in Kalla and Broockman (2020)), lectures, roommates/neighbors and sports teams.

The second broad set of intervention variables are used to characterize the interventions with respect to Allport (1954)'s conditions. The first is *equal status* among groups, with typically members of sports teams or classmates considered equal, but canvassing operations considered as non-equal. The second condition is the clear objective of a *common goal* between participants. For instance, playing with someone from another caste in one's own team in Lowe (2021) is considered as having a common goal, but when the other caste member is on the other team, the common goal condition is not satisfied. The third condition is a *positive contact*. This condition is satisfied if the individual met during the intervention counters the initial stereotype. For instance, in Carrell et al. (2019), African American peers with excellent academic records are considered as inducing a positive contact, but peers with low high-school grades are not. The fourth category is the *support of authorities*. The support of authorities was typically coded as 0 if the focus of the exchange was explicitly not framed as involving the prejudice, e.g. the focus of the course in Scacco and Warren (2018) is to improve computer skills, not inter-religious relations.

The third set of variables relate to other characteristics of the interventions which have been found in the literature to potentially moderate the influence of contact. The first variable from this set is *Personal interaction*, which is equal to zero, for instance, if participants are presented with individuals from the outgroup but do not personally interact with them, as in Grutzeck and Gidycz (1997). The second characteristic is the *friendship potential*, typically defined for members of the same sports teams (Mousa, 2020) but not for teachers (Dessel, 2010). The third variable is the scriptedness of the interactions between members. The intervention was considered as scripted when there is a clear detail of what the participants needed to discuss (Broockman and Kalla, 2016; Freddi et al., 2022), and not scripted if the interaction was more free-form (Barnhardt, 2009; Finseraas et al., 2016). Variables for Allport's conditions, personal interaction, friendship potential and scriptedness were all coded as dummy variables.

Several variables related to the repetition of contacts were also defined as follows. The *number of encounters* with people from the other group was divided into four categories, 0, 1, 2-10 and 10+. The *duration of the contact* (in days) intervention was also categorized as 1 (typically one shot intervention), 2-30 and 30+. The length of *time between intervention and measure* was also coded as 0 (immediately after the intervention, including the end of the year for year-long interventions),

1-30 and 30+ days.

2.3 Analysis methodology

For all papers in the analysis, the effect size was normalized using Cohen (1969)'s d statistic ($d = \frac{Effect\ Size}{Standard\ Deviation}$). The variable was coded so that the effect is positive if contact improves intergroup perceptions (increased trust, more outgroup friends, etc).

The level of analysis was the *outcome*, not the main effect for each paper. This selection was done because, for several papers, different measures are used to capture the effects of contact on different outcomes: immediate effect vs longer-term; individual vs collective level; declared vs behavioral, etc.

The average effects, as well as the heterogeneity measures were performed using the methodology by Deeks et al. (2001): the average effect θ_{IV} is calculated as the weighted average of all treatment effects $\theta_{IV} = \frac{\sum w_i \theta_i}{\sum w_i}$, with reciprocals of the standard errors as weights. The heterogeneity metric is Cochran (1950)'s $Q = \sum w_i (\theta_i - \theta_{IV})^2$.

To investigate which characteristics of the contact matter most for efficacy, a Lasso estimation was performed with the standardized effect as the dependent variable, and all the characteristics presented above as regressors. A post-lasso OLS estimation was then performed.

3 Results

3.1 Descriptive statistics

In Tables 1 and 2 are presented the descriptive statistics of the papers used in the paper. While a significant number of papers were published before 2000, the bulk of the experimental contact literature has been done since 2010, with almost half of papers considered published between 2011 and 2022.

As was highlighted in previous analyses (Lemmer and Wagner, 2015; Paluck et al., 2019), a very large fraction of contact interventions were ran in North America and Western Europe, and no paper was run in Latin America, for instance. A very large fraction of the interventions were also conducted on young samples, from children to university students, although more recent papers focused more on general adult populations. The three modal prejudices studied are race or ethnicity, LGBTQ+ and religion.

One interesting fact is that almost all measures focus on the effects of contact on the entire outgroup, and not the outgroup members specifically concerned. We discuss potential implications in the last section.

Table 1: Descriptive statistics of contexts of papers

Variable	N
Publication year	
≤ 2000	11
2001-2010	17
2011-2022	34
Sample size	
[0,50]	8
[51,100]	6
[101,500]	26
501+	22
Average age	
0-18	16
18-25	28
25+	18
Zone	
Asia	3
Middle-East and Northern Africa	4
North America	35
Oceania	4
Sub-Saharan Africa	5
Western Europe	11
Type of prejudice	
Age	1
Caste	2
Disabilities	6
Gender	2
Immigrants	7
LGBTQ+	10
Police	2
Prisoners	1
Race or ethnicity	25
Religion	6
Type of outcome	
Behavior	21
Explicit beliefs or attitudes	37
Implicit behavior	4
Measure for the entire outgroup	54

Table 2: Descriptive statistics of contact interventions

Variable	N
Army recruits	6
Classmates	9
Collaborative task	2
Discussions	28
Lecture	1
Roommates / neighbors	9
Sports team	7
Allport's conditions	
Equal status	41
Common goal	42
Positive contact	56
Support of authorities	49
Personal interaction	58
Other conditions	
Friendship potential	42
Scriptedness	31
Number of encounters	
0	2
1	16
2-10	16
10+	28
Duration of the contact (in days)	
1	20
2-30	12
30+	30
Length between contact and measure (in days)	
0	32
1-30	19
30+	11

One clear result is that there exists a lot of heterogeneity in the type of contact interventions used. The most common form of intervention typically involves scripted discussions (Broockman and Kalla, 2016; Clochard, 2021). The army, sports teams and roommates also provide special contexts which have been studied a lot. Contacts can last for a long time - e.g. roommates sharing a room for the entire first year of university (Boisjoly et al., 2006; Corno et al., 2019) - or be very short (Page-Gould et al., 2008; Boag and Wilson, 2014).

Contrary to what was found in the broader prejudice-reduction literature (Paluck et al., 2021), the samples for the papers are relatively large, with the median sample consisting of approximately 350 subjects, and while a large fraction of the literature investigated the immediate effects of contact, 11 studies provide tests of enduring effects of contact after one month, with Camargo et al. (2010) measuring the effects of being randomly assigned a Black roommate two years after the end of the first year of university.

3.2 Is contact effective?

The forest plot of effects and standard errors are plotted in Figure 1. We can see that contact is not found to significantly increase prejudice for any paper, and that for the majority of papers, contact induces a significant reduction of prejudice (positive effect).

Meta-analytic results indicate that the average estimated effect of contact is 0.329 (SE = 0.007). This measure is highly significant ($p < 0.01$). This result means, that, on average, contact *is effective at reducing prejudice*. The magnitude is very similar to that found Paluck et al. (2019), and can place, on average, contact as having small to moderate effects, as categorized by Cohen (1969)'s rule of thumb.

There also exists a large heterogeneity between effects ($Q = 4243$, $p < 0.01$). Figure 2 displays the funnel plot of the sample. While we cannot reject a systematic bias, for instance due to publication bias, I do not replicate Paluck et al. (2019)'s result of correlation between effect size and standard error ($p = 0.825$). However, there exists a correlation between the standardized effect and the sample size (Appendix B), which could be a sign of bias. On the other hand, the funnel plot is relatively symmetric, with 30 outcomes lower than the weighted average, and 32 higher. It is therefore arguable that the heterogeneity of observed effects is due to heterogeneity in contexts and methodologies.

3.3 What characteristics matter for contact?

Differences by characteristic In Appendix C are presented effect sizes as a function of the different variables presented in Section 2.

Paper	N	Effect (95% CI)	% Weight
Lowe (Adversarial)	770	-0.35 (-0.75, 0.05)	0.12
Scacco Warren	750	-0.26 (-0.95, 0.42)	0.04
DeVries et al. (exp I)	122	-0.17 (-0.53, 0.20)	0.14
Page-Gould et al	144	-0.07 (-2.70, 2.56)	0.00
Carrell et al. (Low ability)	1602	-0.03 (-2.80, 2.74)	0.00
Kalla Broockman (transgender)	1942	0.05 (0.02, 0.08)	19.55
Clochard	366	0.06 (-0.46, 0.57)	0.07
Kalla Broockman (transgender)	1597	0.06 (0.02, 0.10)	15.47
Kalla Broockman (immigrants)	1505	0.06 (0.01, 0.11)	5.02
Kalla Broockman (transgender)	955	0.07 (0.01, 0.12)	4.90
Kalla Broockman (immigrants)	1354	0.08 (0.02, 0.14)	5.96
Carrell et al. (High ability)	1602	0.09 (-0.07, 0.24)	0.78
Kalla Broockman (transgender)	1044	0.09 (0.02, 0.15)	4.60
Kalla Broockman (immigrants)	1575	0.11 (0.06, 0.16)	7.42
Grutzeck Gdycz	136	0.12 (-0.25, 0.45)	0.14
van Laar et al	2016	0.14 (-0.05, 0.33)	0.56
Lowe (Cooperative)	770	0.15 (0.07, 0.23)	2.54
Finseraas Kotsadam	522	0.19 (0.02, 0.35)	0.69
Mousa	459	0.19 (-0.35, 0.74)	0.06
Clochard et al.	162	0.23 (-0.42, 0.59)	0.05
Dessel	36	0.24 (-0.44, 0.91)	0.04
Broockman Kalla	399	0.24 (0.09, 0.39)	0.80
Hull	43	0.24 (0.02, 0.46)	0.41
Clochard et al.	462	0.25 (-0.09, 0.56)	0.17
Green Wong	54	0.25 (0.01, 0.49)	0.34
Broockman Kalla	429	0.29 (0.15, 0.43)	0.97
Hull	43	0.29 (0.07, 0.52)	0.35
Freddi et al (prior contact)	92	0.30 (-1.66, 2.26)	0.01
Sorensen	727	0.30 (0.15, 0.45)	0.88
Finseraas et al	592	0.31 (-0.00, 0.63)	0.19
Broockman Kalla	285	0.34 (0.19, 0.49)	0.82
Clochard et al.	545	0.35 (0.07, 0.62)	0.25
Broockman Kalla	401	0.35 (0.20, 0.50)	0.92
Clochard	366	0.39 (0.05, 0.72)	0.17
DeVries et al. (exp II)	125	0.39 (-0.01, 0.79)	0.12
Finseraas et al	367	0.39 (-0.06, 0.84)	0.09
Bernhardt	1262	0.40 (0.11, 0.68)	0.23
Sorensen	727	0.42 (0.27, 0.56)	0.88
Cunles-Ross O'meara A	30	0.44 (-0.25, 1.13)	0.04
Markowicz	74	0.45 (-0.01, 0.92)	0.09
Cunles-Ross O'meara A	30	0.53 (-0.19, 1.24)	0.04
Burns Corno La Ferrara	495	0.55 (0.02, 1.07)	0.07
Mousa	459	0.59 (0.19, 0.99)	0.12
Boisjoly et al	1275	0.60 (0.00, 1.20)	0.05
Furuto Furuto	124	0.62 (0.29, 0.95)	0.16
Dahl et al	751	0.68 (0.21, 1.15)	0.09
DeVries et al. (exp III)	195	0.70 (0.42, 0.97)	0.26
Krahe Altwasser	70	0.77 (0.32, 1.22)	0.10
Meshel McGlynn	63	0.93 (0.18, 1.69)	0.03
Yablon	254	0.95 (0.44, 1.46)	0.07
Krahe Altwasser	70	1.03 (0.55, 1.45)	0.10
Camargo et al	255	1.07 (1.03, 1.11)	12.53
Cunles-Ross O'meara B	30	1.14 (0.28, 2.00)	0.03
Marmaros & Sacerdote	1250	1.14 (0.20, 2.09)	0.02
Mousa	459	1.36 (0.68, 2.04)	0.04
Cunles-Ross O'meara B	30	1.37 (0.51, 2.23)	0.03
Camargo et al	351	1.50 (1.44, 1.56)	6.14
Alimo	355	1.52 (1.23, 1.81)	0.23
Camargo et al	351	1.75 (1.26, 2.24)	0.05
Boeng Wilson	142	1.76 (1.49, 2.03)	0.26
Freddi et al (no prior contact)	25	2.78 (1.35, 4.21)	0.01
Camargo et al	255	3.00 (2.72, 3.28)	0.23
Overall, IV	34750	0.33 (0.31, 0.34)	100.00

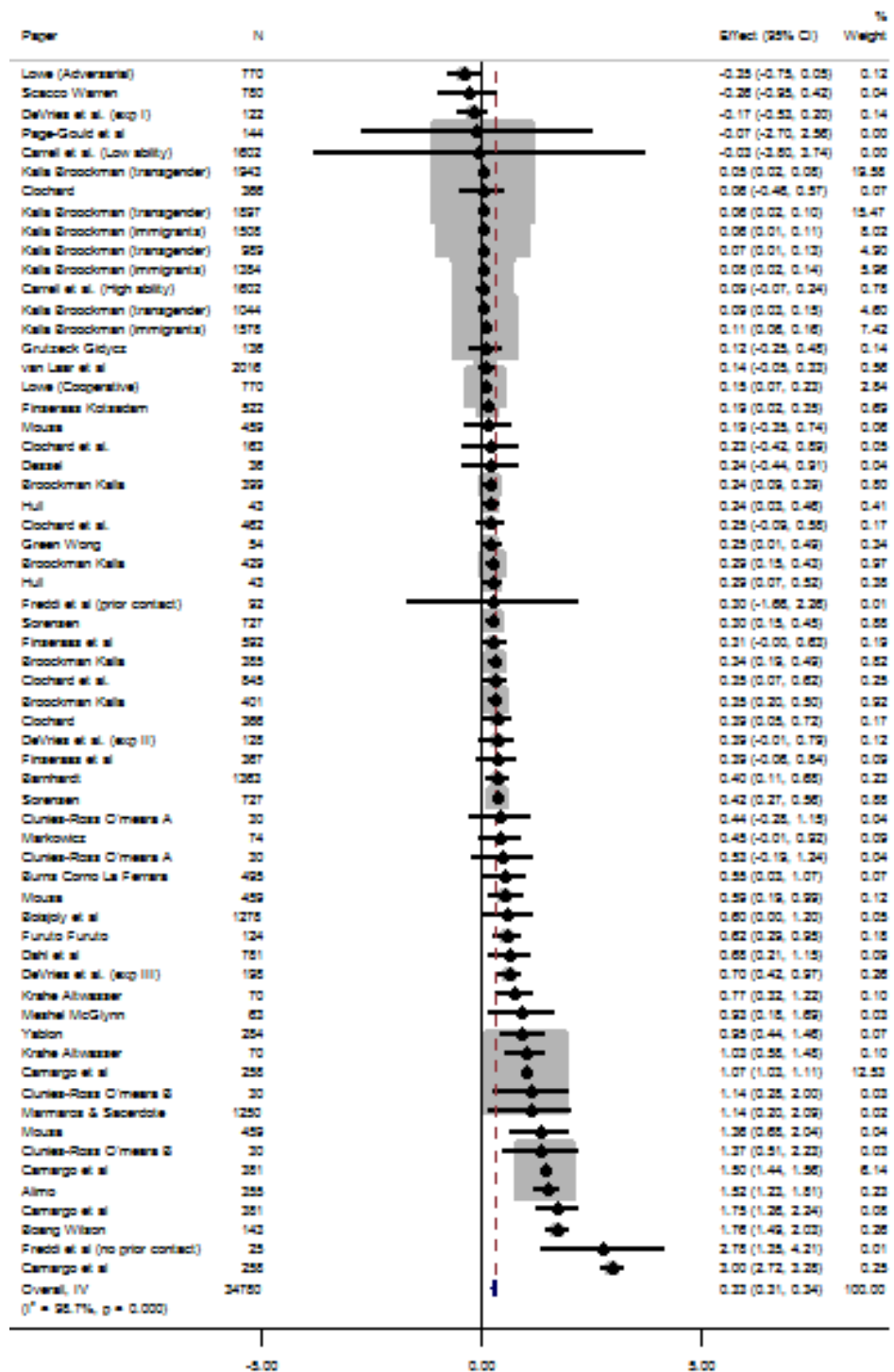$I^2 = 98.7\%$, $p = 0.000$

-5.00   0.00   5.00

Figure 1: Forest plot of estimated treatment effects. Papers are ordered by their point estimates, with the lowest at the top, and the largest treatment effects at the bottom. The overall estimated effect of contact is displayed at the bottom of the graph.
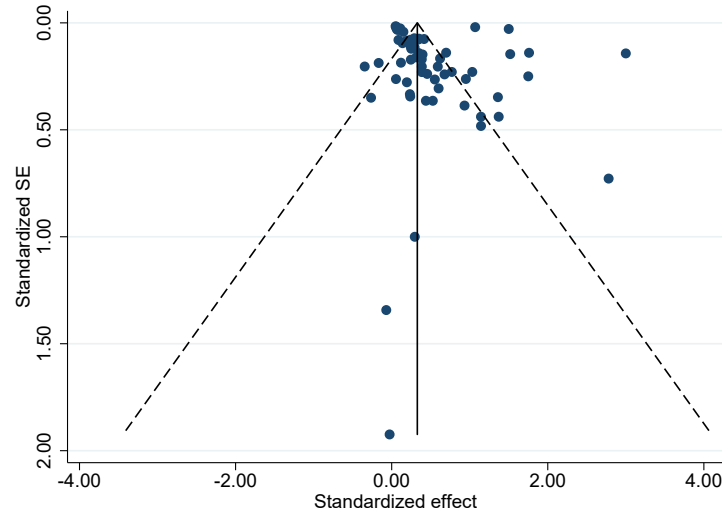
Figure 2: Funnel plot

Age: It appears that studies involving the general population (age 25+) tend to have a lower average effect (Table C.1), and also tend to be have more precise estimates (Figure C.1). This result could be the mere consequence of the fact that studies involving the general adult population tend to be better powered, and therefore provide more accurate estimates, but could also indicate that contact intervention are less effective among adults (although this result would contradict the heterogeneous treatment effects found in Clochard et al. (2022)).

Type of prejudice: No clear pattern emerges as a function of the type of prejudice studied. Most prejudices have a wide range of variation, and although the average effect is higher for papers about prisoners or for race/ethnicity, the number of papers by category is too small to be definitive about significant differences (Table C.2, Figure C.2).

Type of intervention: As for the type of prejudice, no clear pattern emerges from the comparison of the type of contact intervention, in large part due to the low number of observations for the majority of categories (Table C.4, Figure C.4).

Number of encounters: Figure C.5 shows that the average effect is significantly higher for studies with more than 10 encounters. It would thus appear that the higher the number of signals received, the higher the effect.

Duration of contact: The pattern which seems to appear from Figure C.6 and Table C.6 is that the longer the contact, the stronger its effect. This would, as for the number of encounters, point to the result that the more signals received, the more effective the contact.

Time between end of intervention and measurement: No clear pattern emerges from Figure C.7 and Table C.7. One thing clear is that there exist some protocols for which the effect of contact

12

lasts a long time, up to two years following the intervention.

Measure for the entire group: As we can see from Table C.8, and although the sample of measures of the effect of contact toward specific outgroup members is relatively small ($N = 8$), it seems that the effect of contact appears to be larger for the specific individuals met than for the entire outgroup. While the average effect of contact on prejudice is still significantly positive for measures on the effect toward the entire outgroup, it would appear to be only 10% of that of individuals. This can be relatively easily understood as coming from a weaker signal for the entire outgroup than for the specific individuals met.

Lasso   In order to understand which characteristic matter more to the magnitude of the effect of contact on prejudice, I performed a Lasso estimation, using the standardized effect size as the outcome, with all the variables described above as regressors. After the estimation, for the selected variables, I performed an OLS estimation to observe the unbiased effect of these variables on the effect. For the estimations, sample sizes were used as analytical weights.

Results are presented in Table 3. They indicate that the effect of contact on prejudice is stronger if the prejudice considered is Age or Disabilities, and if there is a strictly positive number of encounters.

Interestingly, the algorithm selected neither the characteristics proposed by Allport (1954) nor those proposed by later reviews (Pettigrew and Tropp, 2006; Lemmer and Wagner, 2015; Paluck et al., 2019). It therefore appears that these conditions are neither necessary nor sufficient for effective contact interventions. Of course, this analysis is descriptive, as the conditions are rarely randomly allocated among participating individuals (Lowe (2021) being the exception).

The implications of this meta-analytic work, and hypothetical paths for the future of the contact literature are discussed in the following section.

4   Discussion

In this paper, I conduct a meta-analysis of the literature on the contact hypothesis. While the sample of the initial meta-analysis by Pettigrew and Tropp (2006) consisted almost entirely on descriptive, non-experimental evidence on the effect of contact, the number of experiments using contact is rapidly growing, with the added bonus of widening the geographic origins of samples. While the bulk of the research still takes place in the US, there is now a growing number of studies from other parts of the world, and in particular developing countries.

Moreover, the experimental literature on contact is also becoming more credible, with the use of relatively large sample sizes. Nearly all papers published after 2010 in this analysis would be considered a "large study", using the taxonomy by Paluck et al. (2021), with an average contact

Table 3: Lasso coefficients and Post-Lasso OLS estimation

	Lasso	Post-Lasso OLS
Prejudice = Age	0.408	0.577***
		(0.179)
Prejudice = Disabilities	0.072	0.293
		(0.179)
Number of encounters = 0	-0.070	-0.301*
		(0.179)
Days after end of contact 1-30	-0.016	-0.232
		(0.210)
Constant	0.341	0.588***
		(0.110)
R^2		0.077
No. obs		62
Average dependent variable		0.545
Standard deviation dependent variable		0.644

In column 1 are displayed the Lasso coefficients for the selected variables. Results of the OLS estimation with only the selected coefficients are presented in column 2. * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses.

group of more than 75 individuals.[5] All papers after 2015 have also made use of a pre-analysis plan, reinforcing the credibility of findings.

The present meta-analytic exercise, however, has highlighted three main limitations of the current literature on contact. The first main limitation of the contact literature is the lack of discussion about pre-experimental prejudice. In the typical literature, the absolute level of prejudice is computed from a survey at baseline (before treatment), but no discussion is had about the meaning of observed levels. Moreover, there typically lacks the counterfactual exercise of a comparison group. For instance, White participants are asked whether they agree with the statement "Do you think Black people can be trusted?", but never "Do you think White people can be trusted?". This lack of comparison group forbids much of the literature to identify prejudiced behavior at baseline.

Furthermore, whenever papers do identify pre-experimental levels of discrimination, such as Finseraas et al. (2016), they rarely identify the underlying source of prejudice (i.e. taste-based vs statistical). There also often is a lack of explicit discussion about the absolute levels of prejudice.

The second main limitation of the literature is the lack of consistency in use of the term "contact", as was described in Section 2. The wide variety of protocols falling under the umbrella of contact – from sports leagues to canvassing to interactions between classmates or army recruits

---

[5]The only exception would be the sub-group without minority peers in Freddi et al. (2022).

– makes difficult the comparison of different interventions. In my opinion, future experiments should put more emphasis on the exact content of the contact interventions, not merely the context in which they occur. Now that several meta-analytical works have been carried out, all highlighting the benefits of contact interventions, the focus of the literature should be shifting from the question of *whether* contact interventions are effective, but on the *how* they work. On the note of understanding the mechanisms through which contact can impact prejudice, more effort should be devoted to randomly allocating contact conditions within an experimental sample - *à la* Lowe (2021). This would enable the field to understand which characteristics of protocols are effective at improving intergroup relations. This exercise could help confirm, or disprove, the descriptive results from the lasso analysis above.

The third limitation of the literature is the lack of a theoretical framework which could explain the effects of contact. Of all the papers analyzed in the present paper, only two (Lowe, 2021; Clochard, 2021) make attempts at designing a theoretical framework to explain the effects of contact on prejudice. The literature should work hand in hand with the literature on belief updating. One potential solution could be to integrate a form of updating in the literature on stereotyping (Bordalo et al., 2020). In this regard, analyzing a clear distinction between the effects of contact on the specific individuals met, and the effects of contact on the entire outgroup, could prove useful.

To conclude, the results of the present analysis point to a potential hope for the future of contact interventions. One of the main consistency in the existing literature is the fact that the more interactions participants have with members of the outgroup, the more effective the contact (Page-Gould et al., 2008; Clochard, 2021). This result could mean that a contact functions as a signal about the outgroup, and the higher the number of signals, the larger the shift of the distribution. Because more contacts induce stronger responses, this means that contact can be more suited to reducing statistical discrimination than taste-based discrimination. This could mean that contact is a particularly interesting tool to reduce prejudice in the labor market, as it has been found that the main source of prejudice in this domain is statistical (List, 2004). Now that we know contact interventions can reduce prejudice, efforts should be made to investigate whether contact interventions can also reduce the pernicious effects of prejudice on society and the economy.

References

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003). Fractionalization. *Journal of Economic growth 8*(2), 155–194.

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. W. (1954). *The nature of prejudice*. Addison-wesley Reading, MA.

Barnhardt, S. (2009). Near and dear? evaluating the impact of neighbor diversity on inter-religious attitudes. Technical report, Harvard Working Papers.

Bertrand, M. and E. Duflo (2017). Field experiments on discrimination. In *Handbook of economic field experiments*, Volume 1, pp. 309–393. Elsevier.

Boag, E. M. and D. Wilson (2014). Inside experience: Engagement empathy and prejudice towards prisoners. *Journal of Criminal Psychology 4*(1), 33–43.

Boisjoly, J., G. J. Duncan, M. Kremer, D. M. Levy, and J. Eccles (2006). Empathy or antipathy? the impact of diversity. *American Economic Review 96*(5), 1890–1905.

Bordalo, P., N. Gennaioli, and A. Shleifer (2020). Memory, attention, and choice. *The Quarterly Journal of Economics 135*(3), 1399–1442.

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

Camargo, B., R. Stinebrickner, and T. Stinebrickner (2010). Interracial friendships in college. *Journal of Labor Economics 28*(4), 861–892.

Carrell, S. E., M. Hoekstra, and J. E. West (2015). The impact of intergroup contact on racial attitudes and revealed preferences. Technical report, National Bureau of Economic Research.

Carrell, S. E., M. Hoekstra, and J. E. West (2019). The impact of college diversity on behavior toward minorities. *American Economic Journal: Economic Policy 11*(4), 159–82.

Clochard, G.-J. (2021). Improving the perception of the police by the youth. Technical report, CREST Working Paper.

Clochard, G.-J., G. Hollard, and O. Sene (2022). Low-cost contact interventions can increase inter-ethnic trust when previous contact are scarce: Evidence from senegal. Technical report, CREST Working Paper.

Clunies-Ross, G. and K. O'meara (1989). Changing the attitudes of students towards peers with disabilities. *Australian Psychologist 24*(2), 273–284.

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika 37*(3/4), 256–266.

Cohen, J. (1969). Statistical power analysis for tibe behavioral sciences.

Corno, L., E. La Ferrara, and J. Burns (2019). Interaction, stereotypes and performance: Evidence from south africa. Technical report, IFS Working Papers.

Dahl, G. B., A. Kotsadam, and D.-O. Rooth (2021). Does integration change gender attitudes? the effect of randomly assigning women to traditionally male teams. *The Quarterly Journal of Economics 136*(2), 987–1030.

Danckert, B., P. T. Dinesen, R. Klemmensen, A. S. Nørgaard, D. Stolle, and K. M. Sønderskov (2017). With an open mind: Openness to experience moderates the effect of interethnic encounters on support for immigration. *European Sociological Review 33*(5), 721–733.

Deeks, J. J., D. G. Altman, and M. J. Bradburn (2001). *Statistical Methods for Examining Heterogeneity and Combining Results from Several Studies in Meta-Analysis*, Chapter 15, pp. 285–312. John Wiley & Sons, Ltd.

Deeks, J. J., D. G. Altman, and M. J. Bradburn (2009). *Tolerance and the contact hypothesis: A field experiment*, Chapter 10, pp. 1–23. New York: Oxford University Press.

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

DeVries, D. L., K. J. Edwards, and R. E. Slavin (1977). Biracial learning teams and race relations in the classroom: four field experiments on teams-games-tounament. Technical report, Baltimore: John Hopkins University.

Durante, F. and S. T. Fiske (2017). How social-class stereotypes maintain inequality. *Current opinion in psychology 18*, 43–48.

Finseraas, H., T. Hanson, Å. A. Johnsen, A. Kotsadam, and G. Torsvik (2019). Trust, ethnic diversity, and personal contact: A field experiment. *Journal of Public Economics 173*, 72–84.

Finseraas, H., Å. A. Johnsen, A. Kotsadam, and G. Torsvik (2016). Exposure to female colleagues breaks the glass ceiling—evidence from a combined vignette and field experiment. *European Economic Review 90*, 363–374.

Finseraas, H. and A. Kotsadam (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.

Freddi, E., J. J. J. Potters, and S. Suetens (2022). Brief cooperative contact with ethnic minorities reduces discrimination. Technical report, CEPR Discussion Paper No. DP17098.

Furuto, S. B. and D. M. Furuto (1983). The effects of affective and cognitive treatment on attitude change toward ethnic minority groups. *International Journal of Intercultural Relations 7*(2), 149–165.

Greenwald, A. G., D. E. McGhee, and J. L. Schwartz (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology 74*(6), 1464.

Grutzeck, S. and C. A. Gidycz (1997). The effects of a gay and lesbian speaker panel on college students' attitudes and behaviors: The importance of context effects. *Imagination, Cognition and Personality 17*(1), 65–81.

Hjort, J. (2014). Ethnic divisions and production in firms. *The Quarterly Journal of Economics 129*(4), 1899–1946.

Hull IV, W. F. (1972). Ii. changes in world-mindedness after a cross-cultural sensitivity group experience. *The Journal of Applied Behavioral Science 8*(1), 115–121.

Imperato, C., B. H. Schneider, L. Caricati, Y. Amichai-Hamburger, and T. Mancini (2021). Allport meets internet: A meta-analytical investigation of online intergroup contact and prejudice reduction. *International Journal of Intercultural Relations 81*, 131–141.

Kalla, J. L. and D. E. Broockman (2020). Reducing exclusionary attitudes through interpersonal conversation: Evidence from three field experiments. *American Political Science Review 114*(2), 410–425.

Krahé, B. and C. Altwasser (2006). Changing negative attitudes towards persons with physical disabilities: An experimental intervention. *Journal of Community & Applied Social Psychology 16*(1), 59–69.

Kumar, R., N. Seay, and S. Karabenick (2011). Shades of white: Identity status, stereotypes, prejudice, and xenophobia. *Educational studies 47*(4), 347–378.

Lee, D. and J. Salvatore (2022). Special issue editorial integrating cross-field views on intergroup conflict.

Lemmer, G. and U. Wagner (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.

Lenz, L. and S. Mittlaender (2022). The effect of inter-group contact on discrimination. *Journal of Economic Psychology 89*, 102483.

List, J. A. (2004). The nature and extent of discrimination in the marketplace: Evidence from the field. *The Quarterly Journal of Economics 119*(1), 49–89.

Lowe, M. (2021). Types of contact: A field experiment on collaborative and adversarial caste integration. *American Economic Review 111*(6), 1807–44.

Markowicz, J. A. (2009). *Intergroup contact experience in dialogues on race groups: Does empathy and an informational identity style help explain prejudice reduction?* The Pennsylvania State University.

Marmaros, D. and B. Sacerdote (2006). How do friendships form? *The Quarterly Journal of Economics 121*(1), 79–119.

Meshel, D. S. and R. P. MCGlynn (2004). Intergenerational contact, attitudes, and stereotypes of adolescents and older people. *Educational Gerontology 30*(6), 457–479.

Mousa, S. (2020). Building social cohesion between christians and muslims through soccer in post-isis iraq. *Science 369*(6505), 866–870.

Page-Gould, E., R. Mendoza-Denton, and L. R. Tropp (2008). With a little help from my cross-group friend: Reducing anxiety in intergroup contexts through cross-group friendship. *Journal of personality and social psychology 95*(5), 1080.

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

Paluck, E. L., R. Porat, C. S. Clark, and D. P. Green (2021). Prejudice reduction: Progress and challenges. *Annual review of psychology 72*, 533–560.

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

Rao, G. (2019). Familiarity does not breed contempt: Generosity, discrimination, and diversity in delhi schools. *American Economic Review 109*(3), 774–809.

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

Sorensen, N. A. (2010). *The road to empathy: Dialogic pathways for engaging diversity and improving intergroup relations*. University of Michigan.

Steinmayr, A. (2021). Contact versus exposure: Refugee presence and voting for the far right. *Review of Economics and Statistics 103*(2), 310–327.

Van Laar, C., S. Levin, S. Sinclair, and J. Sidanius (2005). The effect of university roommate contact on ethnic attitudes and behavior. *Journal of Experimental Social Psychology 41*(4), 329–345.

Vertier, P. and M. Viskanic (2018). Dismantling the 'jungle': Migrant relocation and extreme voting in france. Technical report, Center for Economic Studies and ifo Institute.

Whitt, S., R. K. Wilson, and V. Mironova (2021). Inter-group contact and out-group altruism after violence. *Journal of Economic Psychology 86*, 102420.

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.

Appendices

A List of all papers

The description of all the papers with all the variables used in the present paper can be found here.

The list of papers is detailed in Table A.1.

Table A.1: List of all papers used in the analysis

Paper
Alimo (2012)
Barnhardt (2009)
Boag and Wilson (2014)
Boisjoly et al. (2006)
Broockman and Kalla (2016)
Camargo et al. (2010)
Carrell et al. (2019)
Clochard (2021)
Clochard et al. (2022)
Clunies-Ross and O'meara (1989)
Corno et al. (2019)
Dahl et al. (2021)
Dessel (2010)
DeVries et al. (1977)
Finseraas et al. (2016)
Finseraas et al. (2019)
Finseraas and Kotsadam (2017)
Freddi et al. (2022)
Furuto and Furuto (1983)
Deeks et al. (2009)
Grutzeck and Gidycz (1997)
Hull IV (1972)
Kalla and Broockman (2020)
Krahé and Altwasser (2006)
Lowe (2021)
Markowicz (2009)
Marmaros and Sacerdote (2006)
Meshel and MCGlynn (2004)
Mousa (2020)
Page-Gould et al. (2008)
Scacco and Warren (2018)
Sorensen (2010)
Van Laar et al. (2005)
Yablon (2012)

B   Effect size as a function of the sample size
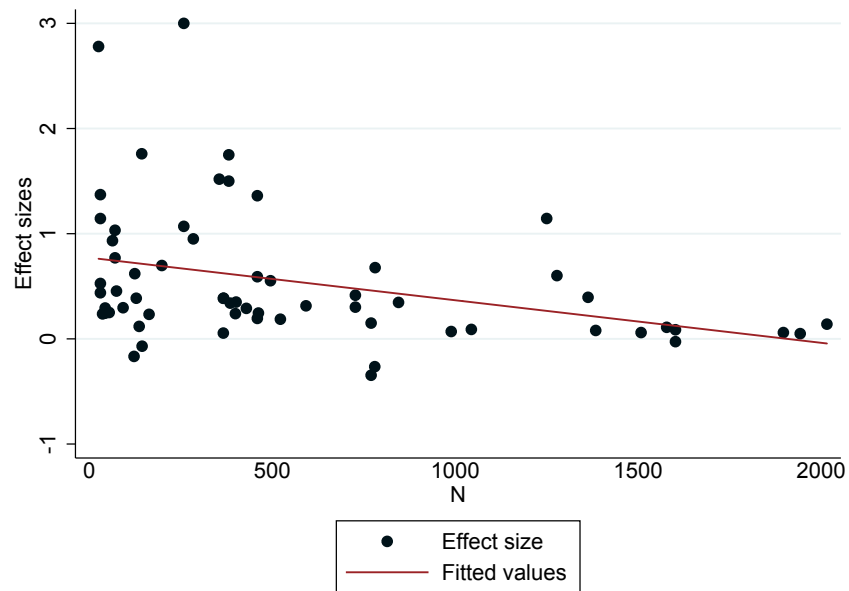
Figure B.1: Effect size as a function of the sample size

Table B.1: Effect size as a function of the sample size

	Effect by sample size
N	-0.0004***
	(0.000)
Constant	0.7718***
	(0.111)
R^2	0.120
No. obs	62

The dependent variable is the standardized effect size. * p<0.10, ** p<0.05, *** p<0.01. Standard errors in parentheses.
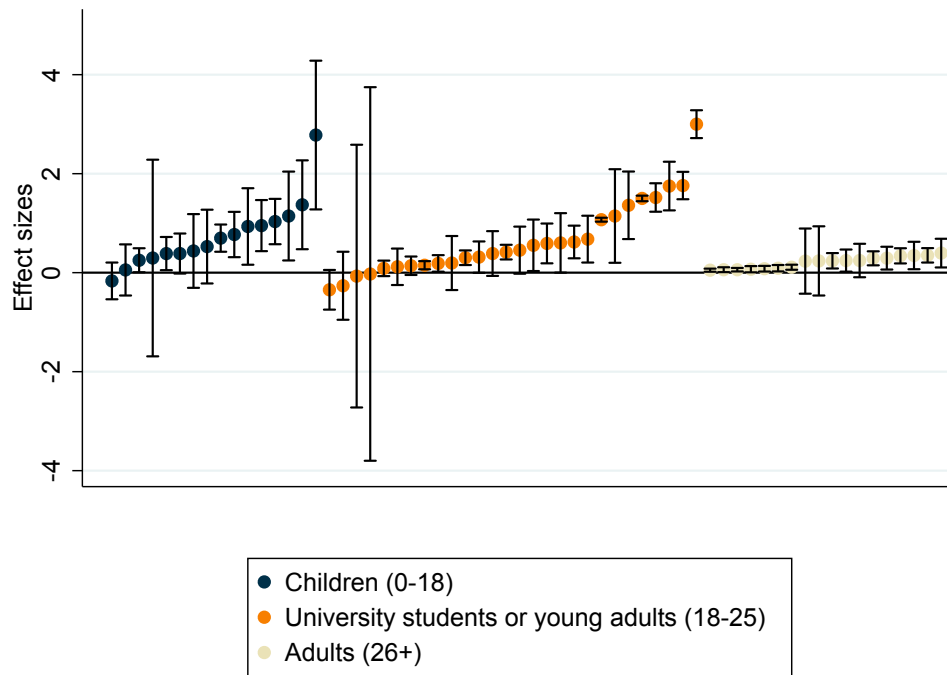
C Effect size as a function of characteristics

Figure C.1: Effect size as a function of age category of the sample

Table C.1: Average effect size per age category

Age category	Average effect	Standard error	Number of measures
0-18	0.492	0.057	16
18-25	0.953	0.014	28
25+	0.084	0.008	18

This table represents average effect sizes and standard errors, from a meta-analysis of all papers in the same age category.
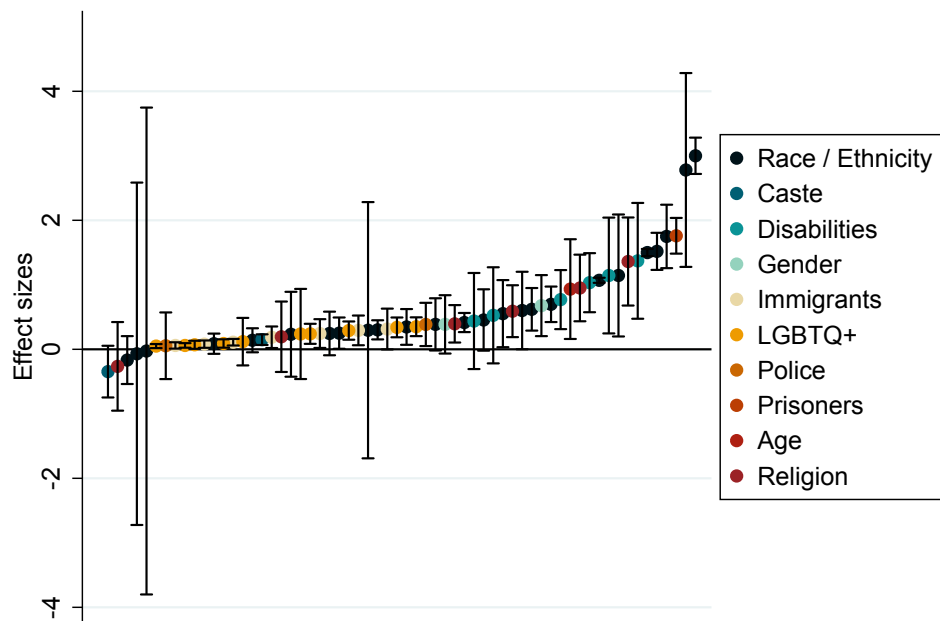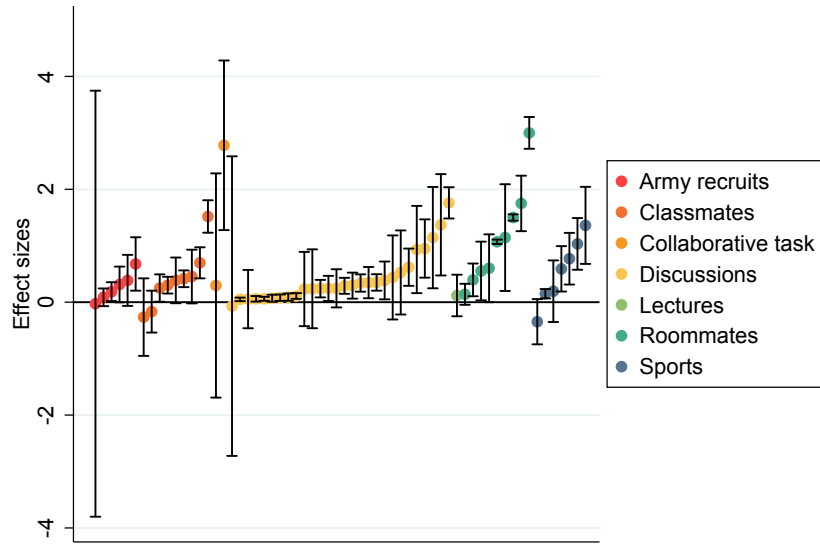
Figure C.2: Effect size as a function of the type of prejudice

Table C.2: Average effect size per prejudice

Type of prejudice	Average effect	Standard error	Number of measures
Age	0.933	0.387	1
Caste	0.130	0.041	2
Disabilities	0.860	0.126	6
Gender	0.525	0.166	2
Immigrants	0.094	0.015	7
LGBTQ+	0.078	0.010	10
Police	0.288	0.143	2
Prisoners	1.760	0.140	1
Race / ethnicity	1.051	0.014	25
Religion	0.508	0.094	6

This table represents average effect sizes and standard errors, from a meta-analysis of all papers in the same category of prejudice.

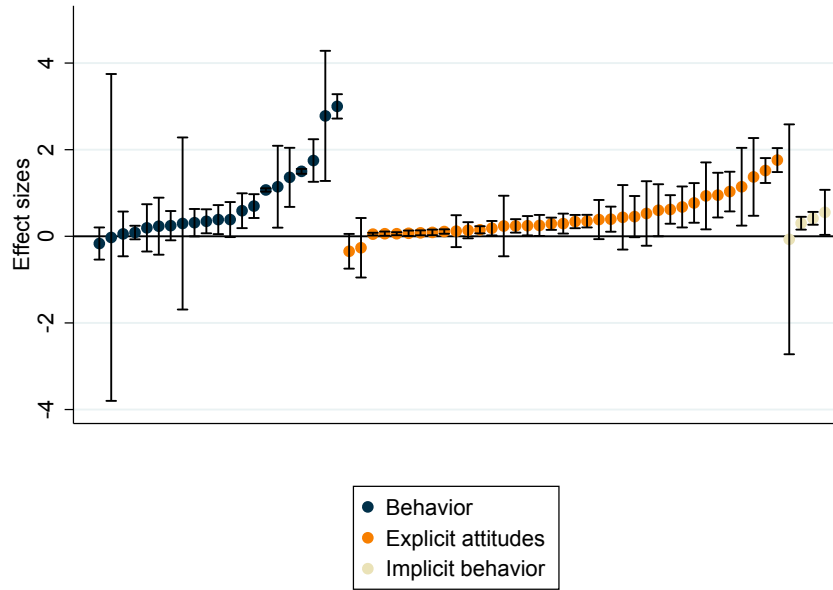Figure C.3: Effect size as a function of the type of intervention

Table C.3: Average effect size per type of contact intervention

Type of contact intervention	Average effect	Standard error	Number of measures
Army recruits	0.191	0.052	6
Classmates	0.437	0.041	9
Collaborative task	1.920	0.589	2
Discussions	0.094	0.008	28
Lecture	0.119	0.186	1
Roommates	1.192	0.016	9
Sports	0.206	0.039	7

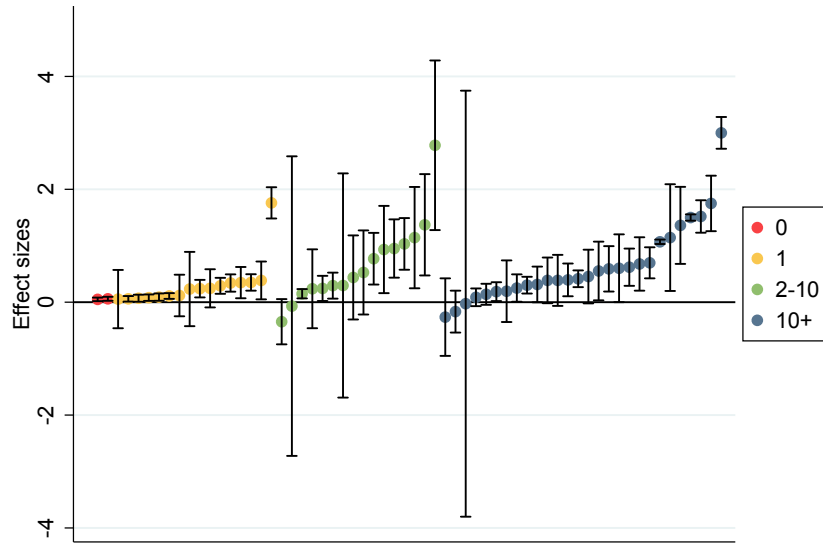This table represents average effect sizes and standard errors, from a meta-analysis of all papers in the same type of contact intervention category.

Figure C.4: Effect size as a function of the type of outcome

Table C.4: Average effect size per type of outcome

Type of outcome	Average effect	Standard error	Number of measures
Behavior	1.129	0.015	21
Explicit attitudes or beliefs	0.104	0.008	37
Implicit behavior	0.365	0.052	4

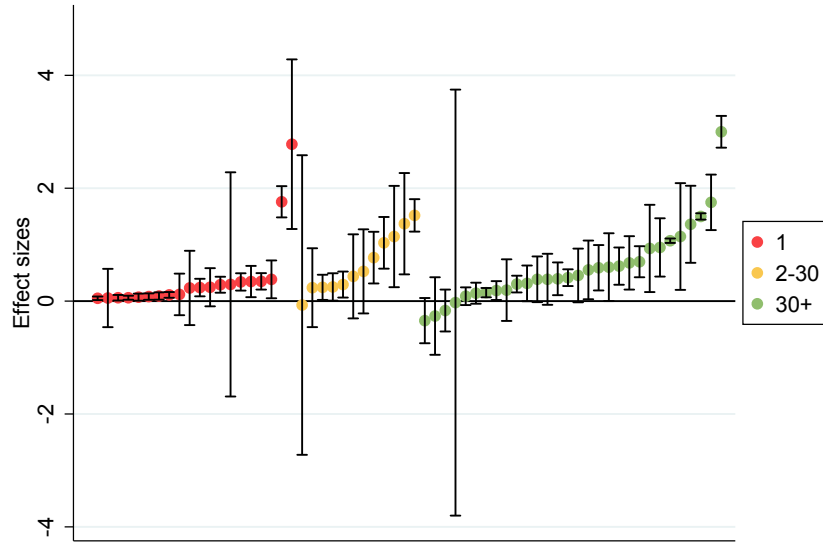This table represents average effect sizes and standard errors, from a meta-analysis of all papers in the same type of outcome category.

Figure C.5: Effect size as a function of the number of encounters

Table C.5: Average effect size per number of encounters

Number of encounters	Average effect	Standard error	Number of measures
0	0.054	0.012	2
1	0.121	0.012	16
2-10	0.238	0.034	16
10+	1.020	0.014	28

This table represents average effect sizes and standard errors, from a meta-analysis of all papers in the same category of number of encounters.

Figure C.6: Effect size as a function of the duration of the contact

Table C.6: Average effect size per duration of the contact intervention

Duration of the contact (days)	Average effect	Standard error	Number of measures
1	0.088	0.008	20
2-30	0.541	0.054	12
30+	0.930	0.013	30

This table represents average effect sizes and standard errors, from a meta-analysis of all papers in the same category intervention duration.
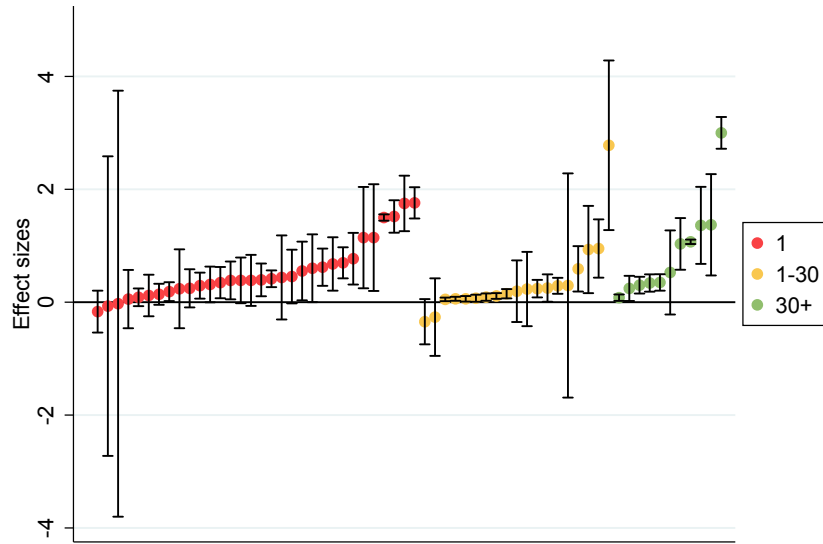
Figure C.7: Effect size as a function of the time between the end of the intervention and the measure

Table C.7: Average effect size per length between the intervention and measure

Length between end of intervention and measure (in days)	Average effect	Standard error	Number of measures
1	0.952	0.020	32
2-30	0.078	0.009	19
30+	0.719	0.015	11

This table represents average effect sizes and standard errors, from a meta-analysis of all papers in the same category length of outcome.
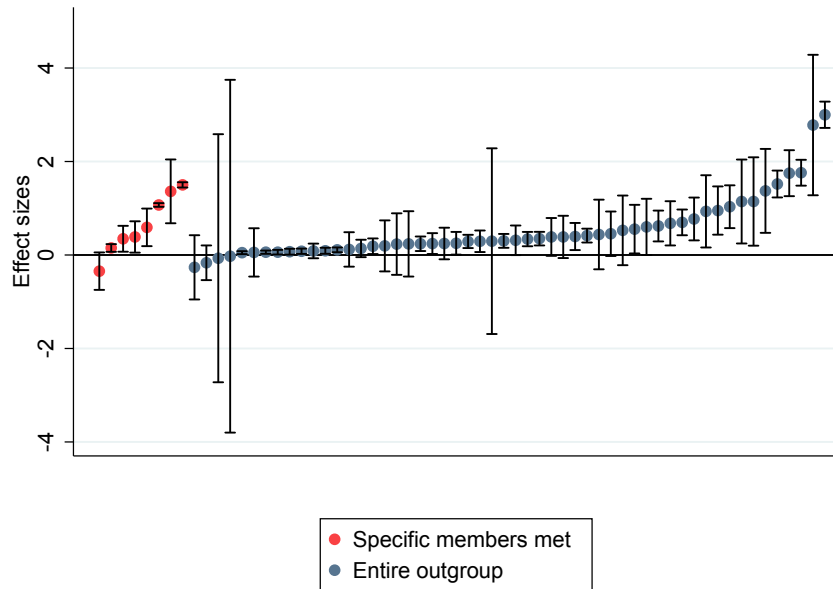
Figure C.8: Effect size as a function of whether the outcome measures prejudice against specific individuals met or the entire outgroup

Table C.8: Average effect size on whether the outcome measures prejudice agains specific individuals met or the entire group

Measure	Average effect	Standard error	Number of measures
Specific members met	1.048	0.015	8
Entire outgroup	0.123	0.008	54

This table represents average effect sizes and standard errors, from a meta-analysis of papers by whether the measure was for the entire outgroup.

CREST
Center for Research in Economics and Statistics
UMR 9194

5 Avenue Henry Le Chatelier
TSA 96642
91764 Palaiseau Cedex
FRANCE

Phone: +33 (0)1 70 26 67 00
Email: info@crest.science
https://crest.science/

The Center for Research in Economics and Statistics (CREST) is a leading French scientific institution for advanced research on quantitative methods applied to the social sciences.

CREST is a joint interdisciplinary unit of research and faculty members of CNRS, ENSAE Paris, ENSAI and the Economics Department of Ecole Polytechnique. Its activities are located physically in the ENSAE Paris building on the Palaiseau campus of Institut Polytechnique de Paris and secondarily on the Ker-Lann campus of ENSAI Rennes.