Can We Really Reduce Ethnic Prejudice Outside the Lab?
A Meta-Analysis of Direct and Indirect Contact Interventions

Gunnar Lemmer and Ulrich Wagner

Philipps-University Marburg, Germany

Author Note

Gunnar Lemmer, Department of Psychology, Philipps-University Marburg; Ulrich Wagner, Department of Psychology and Center of Conflict Studies, Philipps-University of Marburg.

This research was supported by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft e.V., WA 1150/9-1/2) awarded to Ulrich Wagner.

Correspondence concerning this article should be addressed to Gunnar Lemmer, Philipps-University Marburg, Department of Psychology, Gutenbergstrasse 18, 35032 Marburg, Germany. Email: gunnar.lemmer@uni-marburg.de. Telephone: +49(0)6421-28-23632

Word count: 9,258 [including abstract and keywords]

In press: European Journal of Social Psychology

Abstract

The present meta-analysis tested the effectiveness of contact-based interventions for the reduction of ethnic prejudice. Up to now, a meta-analysis summarizing the results of real-world interventions that rest on the intergroup contact theory has been missing. We included evaluations of programs realizing direct (i.e., face-to-face) and/or indirect (i.e., extended or virtual) contact in real-world settings outside the lab. The interventions' effectiveness was tested shortly after their end (k = 123 comparisons, N = 11,371 participants) and with a delay of at least one month (k = 25, N = 1,650). Our data show that contact interventions improve ethnic attitudes. Importantly, changes persist over time. Furthermore, not only direct, but also indirect contact interventions are successful. In addition, contact programs are effective even in the context of a serious societal conflict (e.g., in the Middle East). Although changes are typically larger for ethnic majorities, there is an impact on minorities, too. Finally, contact interventions not only improve attitudes toward individuals involved in the program, their effects also generalize to outgoups as a whole. In sum, social psychology provides an intervention for prejudice reduction that can be successfully implemented in the practical field.

[191 words]

Keywords: meta-analysis, intergroup contact theory, intervention, interethnic relations, ethnic prejudice

Can We Really Reduce Ethnic Prejudice Outside the Lab?

A Meta-Analysis of Direct and Indirect Contact Interventions

Negative relations between members of different ethnic groups (e.g., between immigrants and natives) are still a reality of social life (e.g., European Union Agency for Fundamental Rights, 2014; Hadler, 2012; U.S. Department of Justice, 2013). Ethnic attitudes can be seen as the basis of behavior toward members of other ethnic groups (e.g., Wagner, Christ, & Pettigrew, 2008). Therefore, feasible and effective real-world interventions are needed to reduce ethnic prejudice outside the lab and thereby improve interethnic relations.

Among theories of intergroup relations, the *intergroup contact theory* stands out as a particularly powerful one (see, e.g., Pettigrew & Tropp, 2011). Its basic assumption is that interactions between members of different groups improve attitudes toward the other group(s) and thus reduce intergroup tensions. The contact approach provides a clear and concise guideline for systematic interventions: to improve interethnic relations, persons with different ethnic backgrounds should be brought in direct (i.e., face-to-face) contact or should experience indirect contact (i.e., contact without face-to-face interactions).

However, no quantitative research synthesis exists that focuses on the effects of such *contact interventions* in *real-world settings*. Previous meta-analyses on the intergroup contact theory (Pettigrew & Tropp, 2006; Tropp & Pettigrew, 2005; see also Davies, Tropp, Aron, Pettigrew, & Wright, 2011) have demonstrated the contact-prejudice relation with summaries of mainly correlational studies. The aim of the present research is to extend the literature with several novel facets. First, we want to present the first meta-analysis that focuses on contact interventions conducted outside

the lab. Eighty five percent of the research included in the current research was not considered in Pettigrew & Tropp's (2006) pioneering work. Second, not only information on the short-term, but also on the long-term effects of contact will be summarized. No other review has addressed this issue. Information about the long-term consequences of contact is sorely needed and particularly relevant to understand contact effects in the context of interventions. Third, the current research comprehensively tests and compares both direct and indirect contact programs. To the best of our knowledge, this has not been done before. Fourth, we analyze contact in regions with and without protracted intergroup conflicts. This is a further novel aspect that is not considered in prior meta-analyses. In addition to these unique contributions, another aim of the present paper is to replicate findings of previous reviews concerning reduced contact effects for minorities (Tropp & Pettigrew, 2005) and the generalization of contact effects (Pettigrew & Tropp, 2011) in the context of field interventions where they are of particular interest.

Intergroup Contact Theory

Intergroup contact hypothesis. By holding that direct interactions between members of different groups can reduce prejudice, Allport (1954) formally stated the intergroup contact hypothesis. More than 50 years later, Pettigrew and Tropp (2006) found with a quantitative summary that face-to-face contact is in fact negatively correlated with prejudice (mean r = -.21 at the level of samples). Included was research conducted before 2001. While this finding supports the benefits of direct contact in a general way, it gives no concrete information on the impact of real-world contact interventions and on the long-term effect of contact. Although it is probable that the effects gained from contact are (at least partially) persistent and last when the contact

has ended, existing reviews do not address this very important issue.

Indirect Contact—A theoretical extension of intergroup contact.

The contact hypothesis has been supplemented by several extensions. With the resulting network of hypotheses, the contact approach is now generally seen as an intergroup contact *theory* (see Brown & Hewstone, 2005; Pettigrew & Tropp, 2011). One of the extensions, the consideration of *indirect* forms of intergroup contact, is of special importance for interventions. Several types of indirect contact can be differentiated. The *extended intergroup contact hypothesis* (Wright & Aron, 2010; Wright, Aron, McLaughlin-Volpe, & Ropp, 1997) states that knowing that a member of the ingroup has a close relationship with a member of an outgroup results in improved attitudes toward the respective outgroup. A number of survey studies (e.g., Christ et al., 2010, Study 2; Tausch, Hewstone, Schmid, Hughes, & Cairns, 2011) as well as lab experiments (e.g., Mazziotta, Mummendey, & Wright, 2011; Wright et al., 1997) provide support for this assumption. Second, Amichai-Hamburger and McKenna (2006, see also Amichai-Hamburger, 2012) pointed out the great potential of contact by means of computer technology. Here, we introduce the term *virtual intergroup contact* to characterize indirect contact between members of different groups that do not physically meet but interact via a computer-based communication system. Studies supporting its positive impact already exist (e.g., Schumann, van der Linden, & Klein, 2012; Tavakoli, Hatami, & Thorngate, 2010).

Two further types of indirect contact are reported in the literature but will be not included in the present review. Imagined intergroup contact (Crisp, Stathi, Turner, & Husnu, 2009; Turner, Crisp, & Lambert, 2007) refers to mental simulations of positive intergroup interactions. This type of contact is inherently located in the lab, field

implementations are extremely rare. We will therefore not include imagined contact in the present research and refer to a meta-analysis that specifically deals with it (Miles & Crips, 2014). A further direction sometimes seen as an expansion of intergroup contact is the so-called parasocial contact. Schiappa, Gregg, and Hewes (2005) stated that the exposure to mass media disseminated information about outgroup members is also contact (see also Park, 2012). We do *not* consider this as contact for conceptual reasons: in contrast to direct, extended, and virtual contact, it does not refer to *bidirectional* intergroup interactions but to media-based presentations of outgroup members. Portrayals of close intergroup relations and friendships, on the other hand, can be seen as cases of provided extended contact.

Contact in the context of serious intergroup conflicts. Previous meta-analyses on contact did not consider whether a macro-level or societal conflict is involved. Salomon (2006; see also Wagner & Hewstone, 2012) differentiated between interethnic contact in regions with a recent history of a serious intergroup conflict (e.g., in the Middle East) and contact in regions without it like many places in Europe and North America. Rouhana and Bar-Tal (1998) characterized such conflicts as protracted, central in public life, violent, and perceived as irreconcilable. It can be assumed that contact in these contexts is "surrounded by a general atmosphere of hostility" (Salomon, 2004, p. 262). This also holds true when the actual conflict has already ended, like in Northern Ireland, since it had governed everyday life for a long period of time (e.g., MacGinty, 2010). The current history or even the presence of ongoing intergroup tensions, devaluations, and violence as well as the accompanying collective narratives of the involved groups might create a societal climate with unique challenges for achieving positive effects from contact (see McGary & O'Leary, 1995).

On the other hand, there is also good reason to assume that contact—when organized in a positive way—is effective even in these contexts. One can propose that positive contact experiences are such exceptional and noticeable in conflict areas, that they help to improve attitudes even under these societal conditions (see also Wagner & Hewstone, 2012). Accordingly, survey-based research showed a negative correlation between reported contact and prejudice even in (former) conflict regions like Northern Ireland (e.g., Tausch, Hewstone, Kenworthy, Cairns, & Christ, 2007), South Africa (e.g., Gibson & Claassen, 2010), and Sri Lanka (Malhotra & Liyanage, 2005).

Influence of status: majority and minority groups. Another line of contact research deals with the question whether the strength of the association between contact and ethnic attitudes differs for *ethnic majorities* (e.g., European Americans) and *ethnic minorities* (e.g., African or Asian Americans). Although there is no clear reason why contact should be totally ineffective for minority groups, it can be expected that contact interventions are less effective for minorities. According to Pettigrew and Tropp (2011; Tropp & Pettigrew, 2005**)**, it is likely that contact situations are perceived differently by members of advantaged and disadvantaged groups. Specifically, contact situations are expected to evoke concerns about being confronted with prejudice and discrimination for minority but not for majority members. These concerns, together with a continuous attention to the group's devaluated status, are assumed to reduce but not to eliminate the potential of contact to improve attitudes toward majority groups (see also Stern & West, 2014). Another argument is that minority members' initial attitudes toward majority members could be typically less negative than vice versa (see for e.g. Wagner, Hewstone, & Machleit, 1989). This could be the result of more frequent previous contact experiences (see also Barlow, Hornsey, Thai, Sengupta, & Sibley, 2013) and

could limit the room for improvement for minority groups.

In line with this theorizing, Tropp and Pettigrew (2005) demonstrated with an additional meta-analysis—based on a subset of their review discussed above—that the negative correlation between face-to-face contact and prejudice is larger for ethnic majority samples (mean r = -.24) than for ethnic minority samples (mean r = -.18). A longitudinal study with questionnaire data from multiple countries (Binder et al., 2009) replicated this asymmetric finding.

Generalization of contact effects. To justify contact interventions, it is of particular relevance that their effects generalize to outgroups as a whole. A restricted range would limit the usefulness of interventions dramatically. Some scholars (e.g., Rothbart, 1996) argue that intergroup contact does not have any far-reaching benefits beyond changing interpersonal relations of people involved in the contact situation. In contrast, Pettigrew (1998, 2009) insists that consequences of intergroup contact *generalize*, that is, contact not only changes mutual interpersonal attitudes, but also attitudes toward the target outgroup as a whole and even attitudes toward outgroups not involved in the contact. Pettigrew (2009) and Schmid, Hewstone, Küpper, Zick, & Wagner (2012) as well as Pettigrew and Tropp (2011) provided support for these generalizations in the context of face-to-face contact. Moreover, recent studies showed that generalization effects occur even in the context of indirect contact (Asbrock, Christ, Hewstone, Pettigrew, & Wagner, in prep; Harwood, Paolini, Joyce, Rubin, & Arroyo, 2011). Although the presented findings of non-interventional studies make generalized effects of interventions probable, they do not show it. The generalization of the impact of contact interventions is a further topic of this meta-analysis.

The Present Meta-Analysis

As demonstrated, the intergroup contact theory provides a promising framework for prejudice reduction techniques. But up to now, no meta-analysis exists that comprehensively integrates evaluations of real-world contact interventions. We therefore conducted the first meta-analysis with such a goal.

We use the label *interethnic contact intervention* to refer to programs that: (a) are implemented under naturalistic conditions outside the lab as a (educational) measure to improve intergroup relations in schools, universities, or other applied settings, (b) have the objective to improve interethnic relations by the establishment of direct and/or indirect interethnic contact, and (c) do not just introduce (macro-structural) opportunities for any kind of contact by inducing physical proximity with an ethnically mixed environment, but warrant that direct and/or indirect interethnic contact really occurs as intended. Accordingly, in a study on an interethnic contact intervention, one of the different types of contact (see above and the classification system described in the Method section) is implemented in an applied setting (e.g., classroom) with the intention to improve participants' attitudes toward the groups involved.

Based on the intergroup contact theory and the presented findings, we expected that interethnic contact interventions generally improve ethnic attitudes (Hypothesis 1). We also assumed that they bring about persistent changes (Hypothesis 2). In addition, we expected that both direct and indirect contact interventions improve ethnic attitudes (Hypothesis 3). Furthermore, we hypothesized that contact interventions also reduce ethnic prejudice in settings with a recent history of a severe intergroup conflict and not only in those without such a background (Hypothesis 4). We predicted that contact programs are more effective for ethnic majorities, but are also beneficial for minorities (Hypothesis 5). Finally, we expected (Hypothesis 6) that the effect of contact programs

is not limited to improved attitudes toward individuals involved in the program, but generalizes to outgroups as a whole.

Method

Inclusion Criteria and Exclusion of Studies

The population of studies eligible for inclusion was specified as follows:

1. Independent variable. To be considered, a study had to evaluate the effectiveness of an interethnic contact intervention implemented in a field setting, based on the contact theory, and controlling for the actual occurrence of direct and/or indirect contact (see above). This ruled out the inclusion of macro-level programs (e.g., desegregation initiatives and housing projects) that only open up opportunities for any kind of contact (e.g., Spangenberg & Nel, 1983). For the same reason, we did, for instance, not include student exchange and tourism studies (e.g., Pizam, Fleischer, & Mansfeld, 2002) and evaluations of summer camps with ethnically mixed participants but without structured interethnic encounters (e.g., Amir & Garti, 1977). Additionally, as already mentioned, we did not include studies testing imagined contact and work on parasocial contact.

2. Dependent variable. Relevant studies had to evaluate the impact of a contact intervention with at least one indicator of ethnic prejudice. Appropriate indicators focus on the cognitive (e.g., beliefs or stereotypes about ethnic outgroups), affective (e.g., disliking of ethnic outgroups), and/or on the conative (i.e., associations with intended behavior toward ethnic outgroups) dimension of ethnic attitudes. Moreover, we planned to consider studies that concentrate on actual behavior toward ethnic outgroups. Due to a lack of appropriate research, however, we could not include such evaluations.

3. Evaluation design. We accepted studies with the following evaluation designs:

randomized posttest only with control (POWC), *pretest-posttest with control (PPWC)*, and *pretest-posttest single group (PPSG)*. We did not consider non-randomized studies without a pretest (e.g., Jones, 1994), as potential initial differences between the groups can typically not be adequately controlled. However, we included studies in which a controlled randomized or non-randomized design with a pretest and at least one posttest (PPWC-design) was used. In addition to studies with a randomized POWC-design or PPWC-design, we accepted evaluations with a PPSG-design in which there is no control group but a pretest and at least one posttest. Although such research is susceptible to substantial threats to internal validity (see Shadish, Cook, & Campbell, 2002), we decided to include it since the application of other designs is often not possible in real-world settings. The exclusion of studies with this methodological weaker design would mean to not consider a great deal of relevant results (e.g., Bar-Natan, Rosen, & Salomon, 2010; Connolly, 1992).

4. Available data. Included were studies that allowed calculating an effect size with a sufficient degree of precision, that is, research for which group means, standard deviations, and sample sizes (see equations in Supporting Information A) were available. We also considered studies for which effect sizes could be calculated from other statistics (e.g., *t*-test statistics) by using transformation formulas (e.g., Lipsey & Wilson, 2001). In cases where not enough information was given, the author(s) were contacted—if possible—and we asked for additional data.

5. Language. For pragmatic reasons, we restricted the population of eligible studies to papers that were written in English or German.

Search for Relevant Literature

We used five strategies to search for relevant research. Considered were

documents that were written up to the end of the first half of 2012.

1. Searches in databases. First, we looked for relevant evaluations in databases of multiple scientific disciplines: Psychology (e.g., PsycINFO), Education (e.g., ERIC), Social Sciences (e.g., Sociological Abstracts), Media- and Communication Science (Communication & Mass Media Complete), Sports Science (SPORTDiscus), and Medicine (PubMed). Second, we queried broad multidisciplinary databases (e.g., Google Scholar). Third, to find as many unpublished documents as possible, we searched in international (e.g., NLTD - Networked Digital Library of Theses and Dissertations) and country-specific (e.g., United States: ProQuest Dissertations & Theses) databases for dissertations and master's theses. Fourth, specialized databases containing unpublished literature were searched (e.g., NTIS - National Technical Information Service). Whenever possible, we used multiple search algorithms that combined terms that are characteristic for relevant studies (e.g., contact, attitude, intervention) and a variety of synonyms or related terms (e.g., cooperation, prejudice, workshop) by means of Boolean operators.

2. Manual searches in topic-related journals (e.g., European Journal of Social Psychology, Personality and Social Psychology Bulletin, Journal of Educational Psychology, Journal of Peace Education).

3. Consultation[1] of organizations and experts. A call for papers was repeatedly sent via the listservs of a variety of topic-related scientific organizations from multiple disciplines (e.g., Divisions of the American Psychological Association). Furthermore, we individually contacted scientific experts in the field of study. Third, we wrote to practitioners, evaluators, and organizations connected with interventions to improve interethnic relations.

4. Searches in the proceedings of topic-related conferences (e.g., ISPP annual meetings).

5. Inspection of reference lists. In addition to the aforementioned approaches, we systematically searched the bibliographies of previous reviews—more or less—related to the topic under study (e.g., Aboud et al., 2012; Paluck & Green, 2009; Pettigrew & Tropp, 2006, 2011) as well as the reference list of each potentially relevant document that we found via the other search strategies.

Found documents assessed as relevant (see Results) were subjected to a detailed coding procedure.

Coding Procedure

Study characteristics. A list of the most important study features that we coded can be found in Table 1. With regard to Hypothesis 1 and 2, we registered whether a immediate posttest (less than one month after the intervention) and/or a delayed posttest (one month to one year after the intervention[2]) was conducted. A central coding variable is the type of intervention (see Hypothesis 3). With this paper, we offer a novel classification system. At a general level, we contrasted direct and indirect contact interventions as already described. With more attention to detail, we distinguished two types of direct (contact meetings and cooperative learning programs) and indirect (extended and virtual) contact interventions. *Contact meetings* (e.g., Maoz, 2000) bring together persons with different ethnic backgrounds and in addition explicitly address the relation between the involved groups. This is typically realized by initiating structured intergroup discussions and dialogues. *Cooperative learning programs* (e.g., Cook, 2000) also let persons with different ethnic backgrounds interact. However, participants do not explicitly discuss the existing intergroup relation. Instead, they are requested to

work together on a common learning objective or product that does not relate to interethnic relations. In accordance with the described approaches of indirect contact, we differentiated between two subtypes: extended and virtual contact interventions. *Extended contact interventions* (e.g., Cameron, Rutland, Brown, & Douch, 2006) provide (passages of) books, picture stories, radio plays, or films that explicitly display friendships or positive relations between at least one member of one's own ethnic group and at least one member of an ethnic outgroup. In contrast, *virtual contact interventions* (e.g., Boehm, Kurthen, & Aniola-Jedrzejek, 2008) initiate systematic exchange and discussion activities between members of different ethnic groups by means of computers.

Moreover, to be able to test Hypothesis 4, we registered whether the intervention was realized in a region with a history of a protracted conflict (e.g., in the Middle East, South Africa, Cyprus) or whether this was not the case (e.g., when European Americans were brought in contact with African Americans).

Concerning status (see Hypothesis 5), we classified whether the sample of participants the respective effect is based on (a) had a majority position (e.g., European Americans, when involved in a program together with African Americans), (b) had a minority position (e.g., African Americans, when involved in an intervention together with European Americans), (c) represented a mixture (e.g., when European Americans' and African Americans' data were reported in an aggregate form without differentiating between the two groups), or whether (d) no direct status relation could be inferred (e.g., Americans, when involved in a contact intervention together with Europeans).

Regarding the level of generalization of the dependent variables (see Hypothesis 6), we differentiated between various categories that are described in the Results

section.

In addition to these characteristics that are directly related to our hypotheses, we coded formal features (e.g., publication status), further intervention characteristics (e.g., duration of the program, reason for implementation), as well as the methodological quality (e.g., type of research design).

To specify the interrater reliability, we trained a further coder who coded a random sample of approximately 20% of the included research. Cohen's κ (Cohen, 1960) was above .8 for all variables associated with our hypotheses and can be classified as "almost perfect" (Landis & Koch, 1977).

Effect sizes. We used three effect size indices to display the results of the included studies: Hedges's *g* (for randomized POWC-designs, posttest only with control; see Hedges, 1982; Hedges & Olkin, 1985), Morris's *g* (for PPWC-designs, pretest-posttest with control; see Morris, 2008), and Becker's *g* (for PPSG-designs, pretest-posttest single group; see Becker, 1988). The three indices applied here (see Supporting Information A for details) can be interpreted like Cohen's *d* (difference between an intervention and a control condition expressed in standard deviation units) but consider all the information provided by the design-specific group means and standard deviations as well as a correction factor that prevents effect sizes from small studies from being upwardly biased. Positive values indicate a positive contact effect[3].

Information on the Preliminary Processing of the Effect Sizes

Unit of analysis. Whenever possible, we calculated separate effect sizes for different ethnic groups and age groups within each of the included studies. Since a given study could contribute more than one effect, we used *comparison* (i.e., the comparison between an intervention and a control condition for a specific sample of

participants) instead of study as the unit of our meta-analytic tests. Consequently, for instance, a given study added two comparisons to the meta-analysis when the results were presented separately for European Americans and African Americans.

Clustering of the effect sizes and time of post measurement. Combining effect size estimates from different evaluation designs is a tricky issue. Different effect size indices are used and it is questionable whether the respective findings express the same meaning. Only those designs should be combined in the same analysis that result in equally good estimators of the intervention effect (see Morris & DeShon, 2002). It can be assumed that the findings of evaluations using a randomized POWC-design (posttest only with control) or a PPWC-design (pretest-posttest with control) display the effect of a contact program in a way that it is neither affected by possible pretest discrepancies nor by possible time effects[4] (e.g., differences due to history and/or maturation). Therefore, we analyzed the two designs and the corresponding effect size indices (i.e., Hedges's *g* and Morris's *g*) together in a *primary cluster*.

In contrast, the results of evaluations with a PPSG-design (pretest-posttest single group) reflect both an intervention and a time effect and accordingly provide a systematically less good estimate of the treatment effect. As the nature of these studies is inherently different from those summarized in the primary cluster, we analyzed them and the associated effect size index (i.e., Becker's *g*) in a separate *secondary cluster*. This separation of different designs was also supported by an additional moderator analysis: in a provisional combined meta-analysis of both clusters, studies belonging to the primary cluster differed systematically from those summarized in the secondary cluster ($Q_{model} = 4.04$, $df = 1$, $p < .05$). Consequently, both rational and empirical considerations did not allow a combined integration.

We conducted a further differentiation nested within the primary and secondary cluster. Different time points of post measurement were analyzed with separate meta-analytic integrations. More precisely, within each cluster we differentiated between two sets of analysis: immediate posttest and delayed posttest (see above). This enabled us to make precise tests of the short-term (Hypothesis 1) and long-term (Hypothesis 2) usefulness of contact programs.

In each of the four cluster (primary or secondary) and set (immediate or delayed posttest) combinations that were analyzed separately, we had to eliminate stochastic dependencies between the individual effect sizes (e.g., by averaging across dependent variables within a given comparison; see Supporting Information B).

Meta-Analytic Methods

Meta-analytic models. The expected overall effect of contact interventions (Hypothesis 1 and 2) was tested under the assumptions of the meta-analytic random effects model (REM; see Hedges & Vevea, 1998; Raudenbush, 2009). We used the REM as its multi-level approach allows the individual true effects to vary and to draw unconditional conclusions. At the first level, the observed effect size of each comparison estimates the comparison-specific true effect. At the second level, the individual true effects of all theoretically relevant comparisons (i.e., all existing and future comparisons that satisfy the inclusion criteria) are assumed to be randomly (normally) distributed with a mean that is the average true effect denoted by μ_θ. The variance of this distribution indicates the variability of the true effects (i.e., the heterogeneity) and is denoted by τ^2.

Hypotheses 3 to 6 refer to central study characteristics (e.g., direct vs. indirect contact). We decided to use the meta-analytic mixed effects model (MEM; see

Raudenbush, 2009), which transfers the multi-level approach of the REM to the fixed values of potential moderators. That is, for a given value (e.g., direct contact) of a potential moderator (e.g., direct vs. indirect contact) the assumptions of the REM apply. If a variable is a moderator, the value-specific average true effects differ.

Meta-analytic procedures. We used a restricted maximum likelihood (REML) estimator of τ^2. To gain further insights into the heterogeneity of effects, we employed Cochran's *Q*-Test for homogeneity (Cochran, 1954; Hedges & Olkin, 1985) as well as the I^2-statistic (Higgins & Thompson, 2002). Hypotheses 3 to 6 were tested under the assumptions of the MEM with dummy coded predictor variables and with WLS (weighted least squares) meta-regression models[5] (see Steel & Kammeyer-Mueller, 2002; Viechtbauer, 2008).

For all procedures, we conducted sensitivity analyses to test the robustness of the results. We examined the impact of potential outliers by using externally standardized residuals (Viechtbauer & Cheung, 2010). Moreover, we tested whether an overestimation of the average true effect resulting from a potential publication bias (see Rothstein, Sutton, & Borenstein, 2005) is likely. This was done in three ways: first, we directly controlled whether the effects of published and unpublished documents differ. Second, funnel plots (Light, Singer, & Willett, 1994) that display the individual effect sizes plotted against their corresponding standard errors were inspected. An asymmetric distribution of the effect sizes around the estimated average true effect can signal that the sample of the included research is possibly biased. Third, we statistically tested[6] for funnel plot asymmetry with a rank correlation test (Begg & Mazumdar, 1994) and a regression test (Egger, Davey, Smith, Schneider, & Minder, 1997).

All analyses were conducted with the metafor package (Viechtbauer, 2010) for

R (R Development Core Team, 2010).

Results

Description of the Included Comparisons

By means of the described search strategies and a careful assessment of over 5000 papers of potential relevance, we finally identified 73 documents of relevance for our review. Of these papers, 62 (i.e., 85%) were not considered in the meta-analysis by Pettigrew & Tropp (2006). The 73 identified documents contain 129 independent comparisons ($k = 79$ within the primary cluster, $k = 50$ within the secondary cluster) that satisfy all criteria and that were therefore included in our meta-analytic integrations.

Summary statistics regarding the included comparisons are depicted in Table 1 (see References and Supporting Information C for details about the considered comparisons). Some aspects presented in Table 1 are of particular interest. First, in the primary (34.2%) and the secondary (70.0%) cluster, the time period from 2001 to 2012 is characterized by a greater number of included comparisons than any decade prior to that, indicating that contact interventions are of recent interest. Second, the number of included published and unpublished comparisons is approximately equal. Third, within the secondary cluster a substantial amount of research was conducted in the context of a severe conflict (52.0%), whereas the corresponding percentage is lower within the primary cluster (13.9%). Altogether, the following (former) conflict settings were included: Cyprus, the Middle East (Jewish and Arab-Palestinian persons), Northern Ireland/Republic of Ireland, and South Africa. Fourth, an inspection of the descriptive results concerning the type of contact intervention reveals that most of the included research evaluated direct contact interventions (primary cluster: 79.8%, secondary cluster: 84.0%).

General Effectiveness of Contact Interventions

As stated in Hypothesis 1, we expected a positive impact of real-world contact interventions; with Hypothesis 2 we predicted a long-term effect. Accordingly, we distinguished between immediate (less than one month after the intervention) and delayed (one month up to one year after the intervention) posttests. Moreover, for reasons explained in the Method section, we assessed the effectiveness separately for the primary (i.e., with randomized POWC-designs as well as with PPWC-designs) and secondary (i.e., with PPSG-designs) cluster. Results are depicted in Table 2.

Results for immediate posttests.

Primary cluster. A total of $k = 79$ comparisons with $N = 9{,}212$ participants (intervention: 5,061, control: 4,151) were considered for the meta-analytic integration. The distribution of the individual observed effect sizes is displayed in Figure 1, where they are plotted against their corresponding standard errors. The average true effect was estimated to be $\hat{\mu}_\theta = 0.28$, 95% CI [0.21, 0.35] which indicates more favorable ethnic attitudes in the intervention groups. The null hypothesis, stating that μ_θ is zero, could be rejected ($z = 7.76$, $p < .001$). Cochran's Q-Test suggested heterogeneity ($Q = 251.19$, $df = 78$, $p < .001$), that is, variability among the true effects of the comparisons. The variance of the true effects was estimated to be $\hat{\tau}^2 = 0.06$, 95% CI [0.04, 0.10]. The estimated amount of total variability between the observed effect sizes that is due to heterogeneity was $I^2 = 70.72\%$, 95% CI [58.04, 80.24] and can be classified as "moderate to high" (see Higgins & Thompson, 2002).

We also conducted sensitivity analyses. First, the distribution was inspected with regard to potential outliers. Three[7] of the included comparisons had an absolute externally standardized residual (2.58, 2.33, 2.23) larger than 1.96 and could therefore

be regarded as potential outliers (see Viechtbauer & Cheung, 2010). However, meta-analyses without these cases did not result in substantially different findings. Consequently, the respective comparisons were not excluded. Second, with regard to publication bias, published comparisons and unpublished comparisons did not differ significantly (Q_{model} = 0.23, df = 1, p = .63). Moreover, Figure 1, the rank correlation test (Kendall's τ = 0.06, p = .44), and the regression test (z = 0.08, p = .94) did not suggest funnel plot asymmetry.

Secondary cluster. The meta-analytic integration was based on a total of k = 44 comparisons with N = 2,159 participants. Figure 1 displays the observed effects. The estimated average true effect was $\hat{\mu}_\theta$ = 0.39, 95% CI [0.31, 0.47] and significant (z = 9.19, p < .001). Variability between the individual true effects was detected (Q = 150.85, df = 43, p < .001; $\hat{\tau}^2$ = 0.05, 95% CI [0.02, 0.09]; I^2 = 73.96%, 95% CI [57.60, 83.77]).

Two[8] of the included comparisons had absolute externally standardized residuals (3.16, 2.10) larger than 1.96, but were not discarded as meta-analyses without them showed no substantially different estimates. Published and unpublished comparisons did not differ significantly (Q_{model} = 1.17, df = 1, p = .28). Neither Figure 1 nor statistical tests (Kendall's τ = -0.02, p = .85; regression test: z = -0.78, p = .43) suggested funnel plot asymmetry.

Results for delayed posttests.

Primary cluster. A delayed posttest was conducted for a sample of k = 8 comparisons with N = 1,186 participants (intervention: 617, control: 569). The average of the true effects was estimated to be $\hat{\mu}_\theta$ = 0.23, 95% CI [0.13, 0.34] and differed significantly from zero (z = 4.45, p < .001). Additional analyses indicated absence of

heterogeneity ($Q = 9.11$, $df = 7$, $p = .25$; $\hat{\tau}^2 = 0.003$, 95% CI [0.00, 0.20]; $I^2 = 10.42\%$, 95% CI [0.00, 90.25]).

No potential outliers were detected. A test for differences between published and unpublished comparisons ($Q_{model} = 0.82$, $df = 1$, $p = .36$) as well as tests for funnel plot asymmetry (Kendall's $\tau = 0.50$, $p = .11$; regression test: $z = 1.10$, $p = .27$) indicated that it is unlikely that the results are influenced by a publication bias.

Secondary Cluster. A sample of $k = 17$ comparisons with a total of $N = 464$ participants could be considered. The estimated average of the true effects was $\hat{\mu}_\theta = 0.35$, 95% CI [0.21, 0.50] and significant ($z = 4.68$, $p < .001$). Heterogeneity was uncovered ($Q = 46.27$, $df = 16$, $p < .001$; $\hat{\tau}^2 = 0.05$, 95% CI [0.03, 0.62]; $I^2 = 61.84\%$, 95% CI [40.00, 89.95]).

One comparison[9] had an absolute externally standardized residual (2.58) larger than 1.96. This comparison was, however, not excluded because a meta-analysis without this case had no substantially different results. Published comparisons did not differ significantly ($Q_{model} = 0.83$, $df = 1$, $p = .36$) from unpublished comparisons. Also, the rank correlation test (Kendall's $\tau = 0.00$, $p = 1.0$) as well as the regression test ($z = 0.50$, $p = .61$) did not suggest that the funnel plot is asymmetric.

To summarize, the findings confirm Hypothesis 1 for the primary and the secondary cluster. Interventions based on the intergroup contact theory improve ethnic attitudes. In line with Hypothesis 2, this effect is sustained over time.

A Detailed Look at the Effectiveness of Contact Interventions

To test our Hypotheses 3 to 6, we again examined the primary and secondary cluster separately. Due to the small number of comparisons with a delayed posttest, our analyses were limited to immediate posttests.

Type of contact intervention. As stated in Hypothesis 3, we expected that both direct and indirect contact interventions improve ethnic attitudes. Results can be seen in Table 3.

Within the primary cluster, the average impact of direct programs was estimated to be $\hat{\mu}_\theta = 0.29$ (95% CI [0.21, 0.37], $z = 7.20$, $p < .001$, $k = 63$). Indirect programs had an estimated mean true effect of $\hat{\mu}_\theta = 0.23$ (95% CI [0.07, 0.38], $z = 2.89$, $p < .01$, $k = 16$). The two effects were positive and significantly different from zero, but did not differ from each other ($Q_{model} = 0.47$, $df = 1$, $p = .49$).

In the secondary cluster, direct interventions had an estimated average true effect of $\hat{\mu}_\theta = 0.41$ (95% CI [0.32, 0.50], $z = 8.80$, $p < .001$, $k = 37$), the mean effect of indirect interventions was estimated to be $\hat{\mu}_\theta = 0.33$ (95% CI [0.05, 0.61], $z = 2.89$, $p < .05$, $k = 5$). Again, the impact of each type differed significantly from zero. In addition, two comparisons referred to interventions with direct and indirect components. The average effectiveness of this mixture was estimated to be $\hat{\mu}_\theta = 0.24$ (95% CI [-0.12, 0.60], $z = 1.31$, $p = .19$, $k = 2$). The three types of contact programs (direct, indirect, and their combination) did not differ significantly ($Q_{model} = 0.99$, $df = 2$, $p = .61$).

We also analyzed the data at a more detailed level. As described in the Method section, we designed a system to categorize specific subtypes of contact interventions. The average true effect was estimated to be $\hat{\mu}_\theta = 0.31$ (95% CI [0.21, 0.41], $z = 6.22$, $p < .001$, $k = 39$) for contact meetings, $\hat{\mu}_\theta = 0.25$ (95% CI [0.13, 0.38], $z = 3.96$, $p < .001$, $k = 24$) for cooperative learning methods, $\hat{\mu}_\theta = 0.42$ (95% CI [0.19, 0.65], $z = 3.59$, $p < .001$, $k = 8$) for extended contact programs, and $\hat{\mu}_\theta = 0.08$ (95% CI [-0.13, 0.28], $z = 0.74$, $p = .46$, $k = 8$) for virtual contact interventions. Hence, with the exception of

virtual contact programs, all other subtypes had an estimated effect that differed significantly from zero in the positive direction. Despite the non-significant effect for virtual contact interventions, we found (see Table 3) that the differentiation in different subtypes of contact was not a significant moderator in the primary cluster (Q_{model} = 5.82, df = 3, p = .12).

In reference to the secondary cluster, all of the included evaluations of direct programs tested the impact of contact meetings, all research on indirect contact realized virtual contact, and the two combined cases implemented contact meetings together with virtual contact. Since indirect contact interventions had a significant positive impact in the secondary cluster, this cluster provides some—albeit methodologically weak—evidence for the usefulness of virtual contact programs.

In sum, the findings support Hypothesis 3: direct and indirect contact interventions are effective. When analyzed more closely, a positive impact of contact meetings, cooperative learning methods, and extended contact interventions could be verified. The effectiveness of virtual contact programs was not definitely confirmed, a positive effect of this subtype could only be found in the secondary cluster.

Interventions in the context of serious conflicts. We hypothesized (Hypothesis 4) that contact-based interventions are not only useful in areas free of severe macro-level or societal conflicts, but also in regions with a current or former conflict. Table 3 contains the results of our test.

With regard to the primary cluster, contact interventions that were implemented in the context of a protracted conflict between the involved ethnic groups had an estimated average true effect of $\hat{\mu}_\theta$ = 0.31 (95% CI [0.14, 0.48], z = 3.48, p < .001, k = 11). For other regions, the mean of the true effects was estimated to be $\hat{\mu}_\theta$ = 0.27 (95%

CI [0.19, 0.35], $z = 6.91$, $p < .001$, $k = 68$). Both estimated average true effects were positive and significantly different from zero. The two settings, however, did not differ significantly ($Q_{model} = 0.15$, $df = 1$, $p = .69$).

In reference to the secondary cluster, contact programs conducted in conflict zones had an estimated mean true effect of $\hat{\mu}_\theta = 0.48$ (95% CI [0.36, 0.59], $z = 7.98$, $p < .001$, $k = 20$), interventions that were realized in the absence of a serious conflict had an estimated average effect of $\hat{\mu}_\theta = 0.32$ (95% CI [0.20, 0.43], $z = 5.43$, $p < .001$, $k = 24$). While the two estimated mean true effects were significant, the difference was not ($Q_{model} = 3.75$, $df = 1$, $p = .053$).

Taken together, the findings confirm Hypothesis 4. Contact programs also have a positive influence in settings with high-intensity conflicts between ethnic groups or in those areas that suffered from such a constellation in the recent past.

Ethnic majorities and minorities. We assumed (Hypothesis 5) that contact programs are effective for ethnic majority and minority members, but that intergroup status is a moderator. To test this hypothesis, we considered two groups of comparisons: comparisons that exclusively consisted of data from ethnic majorities and those which only included members of ethnic minorities (see Table 1).

Concerning the primary cluster (see Table 3), contact interventions were more beneficial ($Q_{model} = 4.39$, $df = 1$, $p < .05$[10]) for ethnic majorities ($\hat{\mu}_\theta = 0.38$, 95% CI [0.27, 0.49], $z = 6.84$, $p < .001$, $k = 39$) than for ethnic minorities ($\hat{\mu}_\theta = 0.20$, 95% CI [0.05, 0.35], $z = 2.64$, $p < .01$, $k = 20$). Nonetheless, the estimated average true impact for minority members differed significantly from zero in the positive direction.

In the secondary cluster, comparisons that exclusively included majority members also had a higher estimated average true effect ($\hat{\mu}_\theta = 0.46$, 95% CI [0.29,

0.63], $z = 5.22$, $p < .001$, $k = 12$) than comparisons solely based on minorities ($\hat{\mu}_\theta =$ 0.37, 95% CI [0.22, 0.53], z = 4.65, $p < .001$, $k = 15$). However, the difference was not significant ($Q_{model} = 0.53$, $df = 1$, $p = .47$).

In sum, the findings largely support Hypothesis 5. According to the results for the primary cluster, contact interventions are more effective for ethnic majorities than for ethnic minorities. Nonetheless, the findings for both clusters demonstrate that contact programs also have a positive impact on ethnic minority members.

Generalization of intervention effects. Finally, we expected (Hypothesis 6) a generalization of the impact of contact interventions. The total sample of comparisons can be divided into four categories (see Table 1): (a) exclusively relating to personal relations with individuals included in the study, (b) solely capturing attitudes toward the entire target outgroup, (c) only referring to attitudes toward unspecified ethnic outgroups ("other ethnic groups"), and (d) a combination of the three categories mentioned previously. No comparison exclusively focused on specific outgroups that were not involved in the program. To test Hypothesis 6, comparisons that fit in category a, b, or c were appropriate.

Within the primary cluster (see Table 3), comparisons that only contained measures at the level of individual outgroup members had an average estimated true effect of $\hat{\mu}_\theta = 0.26$ (95% CI [0.08, 0.44], $z = 2.78$, $p < .01$, $k = 12$), comparisons that exclusively measured prejudice at the level of the entire target outgroup had an estimated mean impact of $\hat{\mu}_\theta = 0.32$ (95% CI [0.22, 0.43], $z = 6.00$, $p < .001$, $k = 40$), and, finally, the average true effect of comparisons whose variables were located at the more general level of unspecified ethnic outgroups was estimated to be $\hat{\mu}_\theta = 0.21$ (95% CI [0.05, 0.38], $z = 2.52$, $p < .05$, $k = 16$). There were no significant differences between

the categories (Q_{model} = 1.37, df = 2, p = .51). The estimated average true effects of all three categories differed significantly from zero.

With regard to the secondary cluster, two categories of comparisons were included: comparisons with dependent variables exclusively located at the level of the entire target outgroup ($\hat{\mu}_\theta$ = 0.45, 95% CI [0.34, 0.55], z = 8.50, p < .001, k = 25) and comparisons solely focusing on attitudes toward unspecified outgroups ($\hat{\mu}_\theta$ = 0.33, 95% CI [0.19, 0.47], z = 4.54, p < .001, k = 15). The two estimated mean true effects were significant, whereas the difference between them was not (Q_{model} = 1.67, df = 1, p = .20).

It can be concluded that the effect of contact-based interventions is not restricted to an improvement of personal relations with other individuals. The findings rather demonstrate generalizations to the entire target outgroup and to attitudes toward unspecified ethnic outgroups.

Sensitivity analyses.

Potential outliers within the tested models were examined by using externally standardized residuals. In sum, there were only small numbers of outliers within the models. The results of analyses conducted without these potential outliers did not substantially differ from the reported results. Therefore, the respective comparisons were not eliminated.

Supplementary Results

In addition to the tests of our hypotheses, the influences of further variables were investigated (see Table 1).

In the primary cluster, we found two significant moderators. At first, *type of assignment to conditions* (random vs. non-random) had an influence (Q_{model} = 5.10, df = 1, p < .05). The estimated average true effect of randomized comparisons ($\hat{\mu}_\theta$ = 0.43,

95% CI [0.28, 0.59], $z = 5.61$, $p < .001$, $k = 19$) was higher than the impact of comparisons lacking a random assignment of individuals to conditions ($\hat{\mu}_\theta = 0.24$, 95% CI [0.16, 0.31], $z = 6.19$, $p < .001$, $k = 60$). Second, *mean attrition rate* had a significant impact on the outcome of contact programs (unstandardized $\hat{\beta} = -0.72$, $Q_{model} = 5.91$, $df = 1$, $p < .05$): the effectiveness was higher when loss of participants in the course of the study was lower. The effects of assignment and attrition demonstrate that contact interventions have a larger impact when investigated with research of higher quality. No moderator effect was found for the other coded variables (e.g., duration of contact intervention, age of participants, type of control group) and no moderation was found in the secondary cluster.

Discussion

What Do We Know About the General Effectiveness of Contact Interventions?

We conducted the first meta-analysis on the impact of interethnic contact interventions. The point estimate of the average true effect shortly after the intervention was $\hat{\mu}_\theta = 0.28$ in the primary cluster (i.e., when analyzed with data originating from rigorous evaluation designs) and $\hat{\mu}_\theta = 0.39$ in the secondary cluster (i.e., when pre-post changes were integrated that originate from studies without a control group). The difference between the clusters seems plausible since, conceptually, the comparisons in the primary cluster estimate the intervention effect and the comparisons in the secondary cluster estimate the intervention effect plus a potential time effect (see Morris & DeShon, 2002). For instance, maturation effects as well as the repeated testing with the same instruments together with a lack of a control group for which this is also the case could have artificially enhanced the numerical value of the effect size in the later case. Both average point estimates can be classified as “small” to “medium” (Cohen,

1988) and qualify as "educationally relevant" (Tallmadge, 1977). To illustrate the effect obtained from the data of the primary cluster, on a theoretical 6-point Likert scale with a standard deviation of two, an average participant of a contact intervention is predicted to score about 0.6 scale points (i.e., 28% of the standard deviation) better than an average control group member. That is, when the average control group member would have a score of 3.5, the average intervention participant would have a score of 4.1. With regard to the secondary cluster, the effect would be 0.8 points on the same scale.

Importantly, the described positive impact of interethnic contact interventions is stable over time. No prior review dealt with long-term effects of contact. Our analyses of the delayed posttests that were conducted between one and 12 months after the end of the programs showed estimated true effects similar to those of the direct posttests. Given that the goal of interventions is to generate enduring changes, the findings obtained with our meta-analysis are more encouraging than a large, but rapidly decaying effect would be. However, in none of the included comparisons a measurement of more than 12 months after the end of the program was realized. Thus, further primary research concerning long-time effects is needed.

What Else Do We Know About Contact Interventions?

Indirect contact interventions. Direct contact programs induce face-to-face interactions between members of different ethnicities. In contrast, indirect contact interventions implement contact without physical interactions. Up to now, there has been no comprehensive meta-analysis on indirect contact. As demonstrated by our results, both direct and indirect programs are effective. Moreover, they do not differ regarding their impact.

However, an additional pattern of findings needs to be considered. With

reference to the methodologically better primary cluster, contact interventions were less useful for ethnic minorities as we predicted with Hypothesis 5. Whereas only one of the 16 (i.e., 6.3%) comparisons that model the effect of indirect contact was based on data from a minority group, 19 of the 63 (i.e., 30.2%) comparisons referring to direct contact consisted of data from minorities. This configuration could have masked a difference between the two types of contact programs. For this reason, we additionally tested the influence of the type of contact within the subsample of comparisons that were only based on ethnic majorities. No discrepancy between face-to-face and indirect contact programs was found ($Q_{model} = 0.00$, $df = 1$, $p = .98$). According to our findings, there is no evidence for a differential effectiveness of direct and indirect contact interventions.

A closer look at different subtypes of contact interventions revealed that contact meetings, cooperative learning, and extended contact programs have a positive effect. The results regarding the impact of virtual contact interventions were mixed: our summary of methodologically rigorous research (primary cluster) could not provide support for its effectiveness. In contrast, the findings for the secondary, methodologically weaker cluster showed that virtual contact programs might have positive effects. Based on the results of our meta-analysis, extended contact interventions can be seen as an alternative to direct contact programs (see also Eller, Abrams & Gómez, 2012). This is particularly the case in areas with no or only a few ethnic outgroup members. Furthermore, this variant of contact interventions can typically be realized with less effort and costs as well as under more structured conditions than face-to-face contact interventions. For example, a series of designed stories providing extended contact (see for e.g. Cameron et al, 2006) can be used under structured conditions, repeatedly at will, and—with slight adjustments—in a variety of

different settings (e.g., in schools, colleges and universities).

Contact interventions in conflict regions. Up to now, no meta-analysis has tested the impact of contact in the context of protracted intergroup conflicts. Our results clearly show that interventions based on the intergroup contact theory improve ethnic attitudes in these settings, too. Although some survey-based studies (e.g., Tausch et al., 2007; Gibson & Claassen, 2010) have already signaled that contact can even be effective under the (sometimes) problematic conditions in (former) conflict zones, the findings of the present meta-analysis on interventions are novel and of great relevance. We conclude that programs introducing structured contact between members of different ethnic groups whose recent common history is characterized by reciprocal hostilities, structural inequalities, and severe acts of group-based violence, typically have beneficial effects on the participants' mutual prejudice. Therefore, the implementation of contact interventions is advisable even when their framework conditions are anything but optimal.

Influence of status. In accordance with our prediction and with other research on the intergroup contact theory (Pettigrew & Tropp 2011; Tropp & Pettigrew, 2005), we found a significantly higher effect for ethnic majorities in our methodologically superior primary cluster. In spite of the illustrated discrepancy, the estimated mean effect for ethnic minority members differed significantly from zero both in the primary and in the secondary cluster. Our conclusion is: contact programs are more effective for ethnic majorities, but also have a positive impact on the ethnic attitudes of minority members.

As described earlier, minorities' concerns about being confronted with prejudice and discrimination in contact situations could reduce the potential of contact to improve

attitudes toward majority groups. One can speculate that such concerns are more prominent in the context of physical contact programs than in indirect contact interventions without face-to-face interactions. Unfortunately, due to a lack of appropriate studies focusing on the outcome of indirect contact programs for minority members—we found only one evaluation[11]—we were not able to test this assumption.

Generalization of contact effects. We hypothesized that the impact of contact-based interventions is not restricted to an improvement of interpersonal relations. The findings show that contact programs affect both attitudes toward individuals and attitudes toward the involved outgroup in general. Moreover, our results demonstrate that contact interventions also have a positive outcome when measured at the level of unspecified ethnic outgroups, which is an even stronger indicator of generalized effects of contact interventions. These findings underline the usefulness of contact programs: they are an instrument that can change how societal groups see each other, which, in turn, can set the stage for an improvement of the way they deal with each other.

Directions for Future Research on Intergroup Contact

In addition to the test of our hypotheses, this meta-analysis also revealed some deficits in the available research on intergroup contact and on interventions to improve interethnic relations. We hope that the following package of issues will guide future research.

First, we were able to test whether the effect of contact rapidly fades after the intervention has ended or whether it is sustained over a certain period of time. The available data support the latter. However, only relatively few studies evaluated a prejudice reduction intervention with a delay of one or more months. In addition, no study included a posttest that was conducted more than one year after the end of the

program. Therefore, further primary research investigating long-term effects is needed.

Second, prospective studies on contact interventions also have to investigate the underlying processes. As soon as a substantial number of such studies exist, mediators of the impact of real-world contact programs can be tested meta-analytically. With this we could go a step further and shed light on the question *why* contact *interventions* work. Findings from non-interventional research (Pettigrew & Tropp, 2008) showed that the link between contact and prejudice is mediated by improved empathy and reduced intergroup anxiety. Unfortunately, at the present time, the existing literature does not permit a meta-analytic test of mediators in the context of actual contact interventions.

Third, more studies on indirect contact interventions in real-world settings have to be performed.

Fourth, it has to be clarified if the effectiveness of contact programs can be improved for ethnic minorities. A good starting point would be to systematically evaluate the effectiveness of indirect contact programs for ethnic minorities. If the reduced effect for minority members is really connected with concerns evoked by the contact situation (see Pettigrew & Tropp, 2011; Tropp & Pettigrew 2005), indirect contact interventions might be an alternative that is less obstructive than physical contact, in particular at initial contact stages.

Fifth, upcoming research on contact interventions should also shed light on interindividual differences in the impact of such programs. Survey data have already demonstrated the influence of personality variables like right-wing authoritarianism (Asbrock, Christ, Duckitt, & Sibley, 2012), social dominance orientation (Asbrock et al., 2012), and need for closure (Dhont, Roets, & Van Hiel, 2011). With the available

literature on contact interventions, a meta-analytic investigation of such differential intervention effects is not yet possible.

Sixth, evaluations of contact programs should also focus on changes of actual behavior. Up to now, such work is missing. The findings of future studies would help to gain further insights in the scope of the consequences of contact interventions.

Conclusion

The main message of this quantitative review is that we can reduce ethnic prejudice with real-world contact interventions. Importantly, outcomes are sustained over time. In addition, the positive impact of contact programs is observable for different types of contact interventions, ethnic majorities and minorities, and in contexts with and without a protracted conflict. Moreover, contact programs not only improve attitudes toward individual outgroup members involved in the intervention, but also toward the entire target outgroup and toward unspecified ethnic outgroups.

The meta-analytic results presented here are also of great societal and political importance: if the intention is to reduce prejudice and tensions between societal groups, contact interventions are an effective means. Promising fields for contact interventions are, for example, schools, colleges, universities, work places, and communities. The fact that contact is beneficial even in indirect forms tremendously expands the number of possible realizations of such interventions. In the case of contact programs, social psychology delivers a clearly effective instrument for prejudice reduction.

References

Documents considered in the meta-analytic integrations are marked with an asterisk.

Aboud, F. E, Tredoux, C., Tropp, L. R, Brown, C. S., Niens, U., & Noor, N. M. (2012). Interventions to reduce prejudice and enhance inclusion and respect for ethnic differences in early childhood: A systematic review. *Developmental Review, 32*, 307–336.

*Acevedo Hernandez, J. A. (2012). *Lecturer's awareness and activity engagement with group exercises as tools to develop intercultural sensitivity in students of higher education in University of Twente* (Unpublished master's thesis). University of Twente.

*Al Ramiah, A., & Hewstone, M. (2012).'Rallying around the flag': Can an intergroup contact intervention promote national unity? *British Journal of Social Psychology, 51*, 239–256.

Allport, G. (1954). *The nature of prejudice.* Reading, MA: Addison-Wesley.

Amichai-Hamburger, Y. (2012). Reducing intergroup conflict and promoting intergroup harmony in the digital age. In H, Giles (Ed.), *The Handbook of Intergroup Communication* (pp. 181–193). New York, NY: Routledge.

Amichai-Hamburger, Y., & McKenna, K. Y. A. (2006). The contact hypothesis reconsidered: Interacting via the Internet. *Journal of Computer-Mediated Communication, 11*, 825–843.

Amir, Y., & Garti, C. (1977). Situational and personal influences on attitude change. *International Journal of Intercultural Relations, 1*, 58–75.

Asbrock, F., Christ, O., Duckitt, J., & Sibley, C. G. (2012). Differential effects of intergroup contact for authoritarians and social dominators: A dual process

model perspective. *Personality and Social Psychology Bulletin, 38*, 477–490.

Asbrock, F., Christ, O., Hewstone, M., Pettigrew, T. F., & Wagner, U. (in prep.). *Comparing the secondary transfer effect of direct and extended intergroup contact: The generalization of positive attitudes and its limitations.*

Barlow, F. K., Hornsey, M. J., Thai, M., Sengupta, N. K., & Sibley, C. G. (2013). The wallpaper effect: The contact hypothesis fails for minority group members who live in areas with a high proportion of majority group members. *PLoS ONE*, *8*, e82228.

*Bar-Natan, I., Rosen, Y., & Salomon, G. (2010). *Does friendship between adversaries generalize?* Unpublished manuscript, University of Haifa.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41*, 257–278.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*, 1088–1101.

Binder, J., Zagefka, H., Brown, R., Funke, F., Kessler, T., Mummendey, A., … Leyens, J.P. (2009). Does contact reduce prejudice or does prejudice reduce contact? A longitudinal test of the contact hypothesis among majority and minority groups in three European countries. *Journal of Personality and Social Psychology, 96*, 843–856.

*Boehm, D., Kurthen, H., & Aniola-Jedrzejek, L. (2008). Do international online collaborative learning projects impact ethnocentrism? *E–Learning and Digital Media, 7*, 133–146.

*Boulden, W. T. (2006). Youth leadership, racism, and intergroup dialogue. *Journal of Ethnic & Cultural Diversity in Social Work, 15*, 1–26.

*Bratt, C. (2008). The jigsaw classroom under test: No effect on intergroup relations evident. *Journal of Community & Applied Social Psychology, 18*, 403–419.

Brown, R., & Hewstone, M. (2005). An integrative theory of intergroup contact. *Advances in Experimental Social Psychology, 37*, 255–343.

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

*Cameron, L., Rutland, A., Brown, R. J., & Douch, R. (2006). Changing children's intergroup attitudes towards refugees: Testing different models of extended contact. *Child Development, 77*, 1208–1219.

*Cernat, V. (2011). Extended contact effects: Is exposure to positive outgroup exemplars sufficient or is interaction with ingroup members necessary? *Internet and Higher Education, 14*, 251–261.

Christ, O., Hewstone, M., Tausch, N., Wagner, U., Voci, A., Hughes, J., & Cairns, E. (2010). Direct contact as moderator of extended contact effects. *Personality and Social Psychology Bulletin. 36*, 1662–1674.

*Clark, M. A. (1998). *The effects of a cross-cultural mentoring program and diversity training on preservice teachers' attitudes and beliefs regarding diverse students* (Unpublished doctoral dissertation). University of Florida.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics, 10*, 101–129.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46.

Cohen. J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Connolly, E. (1992). *The Edu-Train project and the reduction of prejudice* (Unpublished master's thesis). Rand Afrikaans University, Johannesburg.

*Connolly, E. (1992). *The Edu-Train project and the reduction of prejudice* (Unpublished master's thesis). Rand Afrikaans University, Johannesburg.

Cook, J. E. (2000). *Effect of cooperative learning (student teams-achievement divisions) on African American and Caucasian students' interracial friendships* (Unpublished doctoral dissertation). Louisiana Tech University.

*Cook, J. E. (2000). *Effect of cooperative learning (student teams-achievement divisions) on African American and Caucasian students' interracial friendships* (Unpublished doctoral dissertation). Louisiana Tech University.

*Cookston, R. R. (1973). *Effects of a short-term, intensive, interracial living experience on interracial social distance and attitudes toward interracial issues* (Unpublished doctoral dissertation). East Texas State University.

Crisp, R. J., Stathi, S., Turner, R. N., & Husnu, S. (2009). Imagined intergroup contact: Theory, paradigm, and practice. *Social and Personality Psychology Compass, 2*, 1–8.

Davies, K., Tropp, L. R., Aron, A., Pettigrew, T. F., & Wright, S. C. (2011). Cross-group friendships and intergroup attitudes: A meta-analytic review. *Personality and Social Psychology Review, 15*, 332–351.

*Derbaum, M. C. (1981). *The effects of multi-ethnic children's stories on the attitudes of White children in grades K-1 and 2-3* (Unpublished doctoral dissertation). University of Pittsburgh.

Dhont, K., Roets, A., & Van Hiel, A. (2011). Opening closed minds: The combined effects of intergroup contact and need for closure on prejudice. *Personality and Social Psychology Bulletin, 37*, 514–528.

*Dodson, J. P. (1970). *Participation in a bi-racial encounter group: Its relationship to acceptance of self and others, racial attitudes, and interpersonal orientations* (Unpublished doctoral dissertation). Purdue University.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629–634.

Eller, A, Abrams, D., & Gómez, Á. (2012). When the direct route is blocked: The extended contact pathway to improving intergroup relations. *International Journal of Intercultural Relations, 36*, 637–646.

*Ertmer, P. A., Newby, T. J., Yu, J. H., Liu, W., Tomory, A., Lee, Y. M., Sendurur, E., & Sendurur, P. (2011). Facilitating students' global perspectives: Collaborating with international partners using Web 2.0 technologies. *Internet and Higher Education, 14*, 251–261.

European Union Agency for Fundamental Rights (2014). *Annual report on fundamental rights*. Retrieved from fra.europa.eu/sites/default/files/fra-2014-annual-report-2013-0_en.pdf

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

*Ganayem, A., Shonfeld, M., Hoter, E., & Walter, J. B. (2010, September). *A longitudinal field experiment among Israeli Jews and Arabs.* Paper presented at the INTER Network Conference, Malta.

*Gant, J. L. (1971). *The development and evaluation of a short-term human relations training program to prepare prospective teachers for multi-racial classrooms* (Unpublished doctoral dissertation). The Florida State University.

*Gardner, R. C., Kirby, D. M., Smythe, P. C., Dumas, G., Zelman, M., & Bramwell, J. R. (1974). Bicultural excursion programs: Their effects on students' stereotypes, attitudes and motivation. *The Alberta Journal of Educational Research, 20*, 270–277.

Gibson, J. L., & Claassen, C. (2010). Racial reconciliation in South Africa: Interracial contact and changes over time. *Journal of Social Issues, 66*, 255–272.

*Gonzalez, A. (1979). *Classroom cooperation and ethnic balance* (Unpublished doctoral dissertation). University of California, Santa Cruz.

*Grant, A. (2006). *The development of global awareness in elementary students through participation in an online cross-cultural project* (Unpublished doctoral dissertation). Louisiana State University and Agricultural & Mechanical College.

*Green, D. P., & Wong, J. S. (2009). Tolerance and the contact hypothesis: A field experiment. In E. Borgida (ed.), *The political psychology of democratic citizenship* (pp. 228–246). New York, NY: Oxford University Press.

Hadler, M. (2012). The influence of world societal forces on social tolerance. A time comparative study of prejudices in 32 countries. *The Sociological Quarterly, 53***,** 211–237.

Harwood, J., Paolini, S., Joyce, N., Rubin, M., & Arroyo, A. (2011). Secondary Transfer Effects from Imagined Contact. *British Journal of Social Psychology, 50*, 180–189.

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

Academic Press.

Hedges, L. V, & Vevea, J. L (1998). Fixed- and Random-Effects Models in Meta-Analysis. *Psychological Methods, 3*, 486–504.

Hedges, L. V. (1982). Estimation of effect size from a series of independent studies. *Psychological Bulletin, 92*, 490–499.

*Hertz-Lazarowitz, R., Kuppermintz, H., & Lang, J. (1998). Arab-Jewish coexistence: Beit Hagafen coexistence programs. In E. Weiner (Ed.), *The handbook of interethnic coexistence* (pp. 565–584). New York, NY: Continuum.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539–1558.

*Hoffman, A. J, Wallach, J., & Sanchez, E. (2009). Reducing ethnocentric ideology via multiethnic community service work: Planting Seeds of Hope. *Making connections: Interdisciplinary Approaches to Cultural Diversity, 11*, 40–50.

*Horenczyk, G., & Bekerman, Z. (1997). The effects of intercultural acquaintance and structured intergroup interaction on ingroup, outgroup, and reflected ingroup stereotypes. *International Journal of Intercultural Relations, 21*, 71–83.

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

*Jackson, R., Jr. (1998). *The effects of cooperative learning on the development of cross-racial friendships* (Unpublished doctoral dissertation). Lehigh University.

*Jin, L., & Erben, T. (2007). Intercultural learning via instant messenger interaction. *CALICO Journal, 24*, 291–311.

*Johnson, D. W., & Johnson, R. T. (1982). Effects of cooperative, competitive, and individualistic learning experiences on cross-ethnic interaction and friendships.

The Journal of Social Psychology,118, 47–58.

Jones, D. R. (1994). *Effects of cooperative learning on student behavior, attendance, and cross-ethnic relationships.* (Unpublished doctoral dissertation). University of Kansas.

*Kowalski, K. L. (1998). The impact of vicarious exposure to diversity on preschoolers' emerging ethnic/racial attitudes. *Early Child Development and Care, 146*, 41–51.

*Kropiunigg, U., & Pabst, B. (2007). Selbstwertsteigerung und Vorurteilsreduktion bei Jugendlichen eines multiethnischen Peace Camps. *Psychotherapie Forum, 15*, 63–72.

*Küchel, J., & Beelmann, A. (2008, July). *Effects of an extended contact intervention on preventing prejudice in elementary school children*. Poster presented at the International Congress of Psychology, Berlin, Germany.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 1159–1174.

*Lazovsky, R. (2007). Educating Jewish and Arab children for tolerance and coexistence in a situation of ongoing conflict: An encounter program. *Cambridge Journal of Education, 37*, 391–408.

Light, R. J., Singer, J. D., & Willett, J. B. (1994). Displaying and communicating findings from a meta-analysis. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 439–453). New York, NY: The Russell Sage Foundation.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis (Applied Social Research*

*London, L. H. (1995). *Kids' College: Enhancing children's understanding and*

acceptance of cultural diversity (Unpublished doctoral dissertation). University of South Carolina.

*London, L. H., Tierney, G., Buhin, L., Greco, D. M., & Cooper, C. J. (2002). Kids' College: Enhancing children's appreciation and acceptance of cultural diversity. *Journal of Prevention and Intervention in the Community, 24*, 63–78.

*Luiz, D., & Krige, P (1985). The effect of social contact between South African White and Colored adolescent girls: A follow-up study. *The Journal of Social Psychology, 125*, 407–408.

*Lyras, A. (2007). *Characteristics and psycho-social impacts of an interethnic educational sport initiative on Greek and Turkish Cypriot youth* (Unpublished doctoral dissertation). University of Connecticut.

MacGinty, R. (2010). No war, no peace: Why so many peace processes fail to deliver peace. *International Politics, 47*, 145–162.

Malhotra, D., & Liyanage, S. (2005). Long-term effects of peace workshops in protracted conflicts. *Journal of Education, 37*, 391–408.

*Mania, E. W., Riek, B. M., Gaertner, S. L., McDonald, S. A., & Lamoreaux, M. J. (2008). *Quantitative evaluation of the Building Bridges for Peace program*. Unpublished manuscript.

Maoz, I. (2000). An experiment in peace: Reconciliation-aimed workshops of workshops of Jewish-Israeli and Palestinian youth. *Journal of Peace Research, 37*, 721–736.

*Maoz, I. (2000). An experiment in peace: Reconciliation-aimed workshops of workshops of Jewish-Israeli and Palestinian youth. *Journal of Peace Research, 37*, 721–736.

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

Mazziotta, A., Mummendey, A., & Wright, S. C. (2011). Vicarious intergroup contact effects: Applying social-cognitive theory to intergroup contact research [Special Issue]. *Group Processes and Intergroup Relations, 14*, 255-274,.

McGarry, J., & O'Leary, B. (1995). *Explaining Northern Ireland: Broken images.* Oxford, UK: Blackwell.

Methods Series, Vol. 49). Thousand Oaks, CA: Sage.

Miles, E., & Crisp, R. J. (2014). A meta-analytic test of the imagined contact hypothesis. *Group Processes Intergroup Relations*. *17*, 13–26.

*Miller, K. M. (1994). *The effects of interethnic contact in cooperative learning groups on student attitudes toward race and ethnicity* (Unpublished doctoral dissertation). University of San Francisco.

Morris, S. (2008). Estimating effect sizes from pretest-posttest-control group Designs. *Organizational Research Methods, 11*, 364–386.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*, 105–125.

*Nganga, R. W. (2006). *Impact of cross-cultural interaction on counselor trainees' development of cultural empathy and intercultural sensitivity* (Unpublished doctoral dissertation). University of Wyoming.

*O'Connor, J., Ruddle, H., & Keenaghan, C. (1992). *A Model of managed co-operation:*

An evaluation of Co-Operation North's School and Youth Links Scheme. Dublin: Co-Operation North.

*Ohm, R. M. (1987). *Constructing and reconstructing social distance attitudes* (Unpublished doctoral dissertation). Arizona State University.

*Oishi, S. S. (1983). *Effects of team-assisted individualization in mathematics on cross-race and cross-sex interactions of elementary school children* (Unpublished doctoral dissertation). University of Maryland College Park.

*Oishi, S., Slavin, R. E., & Madden, M. A. (1983, April). *Effects of student teams and individualized instruction on cross-race and cross-sex friendship*s. Paper presented at the Annual Convention of the American Educational Research Association, Montreal, Quebec, Canada.

*O'Neill, E. J. (2008). *Intercultural competence development: Implementing International Virtual Elementary Classroom Activities into public schools in the U.S. and Korea* (Unpublished doctoral dissertation). University of Virginia.

*Otis, M. D., & Loeffler, D. N. (2005). Changing youths' attitudes toward difference: A community-based model that works. *Social Work with Groups, 28*, 41–64.

Paluck, E. L., & Green, D. P. (2009). Prejudice Reduction: What works? A critical look at evidence from the field and the laboratory. *Annual Review of Psychology, 60*, 339–367.

Park, S.Y. (2012). Mediated intergroup contact: Concept explication, synthesis, and application. *Mass Communication and Society, 15*, 136–159.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology, 49*, 65–85.

Pettigrew, T. F. (2009). Contact's secondary transfer effect: Do intergroup contact

effects spread to non-participating outgroups? *Social Psychology, 40*, 55–65.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90*, 751–783.

Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology*, *38*, 922–934

Pettigrew, T. F., & Tropp, L .R. (2011). *When groups meet: The dynamics of intergroup contact*. London, UK: Routledge/Psychology Press.

Pizam, A., Fleischer, A., & Mansfield, Y. (2002). Tourism and social change: The case of Israeli ecotourists visiting Jordan. *Journal of Travel Research, 41*, 177–184.

R Development Core Team (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria.

Raudenbush, S. W. (2009). Analyzing effect sizes: Random effects models. In H. Cooper, L. V. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 295–316). New York, NY: The Russell Sage Foundation.

Rothbart, M. (1996). Category-exemplar dynamics and stereotype change. *International Journal of Intercultural Relations, 20*, 305–321.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta analysis: Prevention, assessment and adjustments*. Chichester, UK: Wiley.

Rouhana, N. N., & Bar-Tal, D. (1998). Psychological dynamics of intractable ethnonational conflicts. The Israeli–Palestinian case. *American Psychologist, 53*, 761–770.

*Ryan, M. M. (1975). *The effects of a New Brunswick, Canada studies project upon the*

attitudes of elementary students toward native Indians and themselves (Unpublished doctoral dissertation). The University of New Brunswick, Canada.

Salomon, G. (2004). Does peace education make a difference in the context of an intractable conflict? *Peace and Conflict: Journal of Peace Psychology, 10*, 257–274.

Salomon, G. (2006). Does peace education really make a difference? *Peace and Conflict: Journal of Peace Psychology, 12*, 37–48.

*Sayler, R. I. (1969). *An exploration of race prejudice in college students and interracial contact* (Unpublished doctoral dissertation). University of Washington.

Schiappa, E., Gregg, P. B., & Hewes, D. E. (2005). The parasocial contact hypothesis. *Communication Monographs, 72*, 92–115.

*Schleien, S. M. (2008). *Confronting the intractable: An evaluation of the Seeds of Peace experience* (Unpublished doctoral dissertation). University of Waterloo, Canada.

Schmid, K., Hewstone, M., Küpper, B., Zick, A., & Wagner, U. (2012). Secondary transfer effects of intergroup contact: A cross-national comparison in Europe. *Social Psychology Quarterly*, *75*, 28–51.

*Schuitema, J., & Veugelers, W. (2011). Multicultural contacts in education: a case study of an exchange project between different ethnic groups. *Educational Studies*, *37*, 101–114.

Schumann, S. van der Linden, N., & Klein, O. (2012). Bridging the Gap on Facebook: Assessing Intergroup Contact and Its Effects for Intergroup Relations. *Cyberpsychology, Behavior, and Social Networking, 15*, 411–416.

*Seaman, J., Beightol, J., Shirilla, P., & Crawford, B. (2009). Contact theory as a framework for experiential activities as diversity education: An exploratory study. *Journal of Experiential Education, 32*, 207–225.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.

*Sharan, S., Kussell, P., Hertz–Lazarowitz, R., Bejarano, Y., Raviv, S., & Sharan, Y. (1984). *Cooperative learning in the classroom: Research in desegregated schools*. Hillsdale, NJ: Lawrence Erlbaum.

*Slavin, R. E. (1979). Effects of biracial learning teams on cross-racial friendships. *Journal of Educational Psychology, 71*, 381–387.

*Slavin, R. E., & Oickle, E. (1981). Effects of cooperative learning teams on student achievement and race relations: Treatment by race interactions. *Sociology of Education, 54*, 174–180.

*Slone, M., Tarrasch, R., & Hallis, D. (2000). Ethnic stereotypic attitudes among Israeli children: Two intervention programs. *Merrill-Palmer Quarterly, 46*, 370–389.

*Smith, F. T. (1943). *An experiment in modifying attitudes toward the Negro.* New York, NY: Columbia University.

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

Spangenberg, J., & Nel, E. M. (1983). The effect of equal-status contact on ethnic attitudes. *Journal of Social Psychology, 121*, 173–180.

Steel, P. D., & Kammeyer-Mueller, J. D. (2002). Comparing meta-analytic moderator

estimation techniques under realistic conditions. *Journal of Applied Psychology, 87*, 96–111.

Stern, C., & West, T. V. (2014) Circumventing anxiety during interpersonal encounters to promote interest in contact: An implementation intention approach. *Journal of Experimental Social Psychology*, 50, 82–93.

*Tackaberry, S. N. (1980). *Cooperation versus competition: Effects on Hispanic and Anglo children* (Unpublished doctoral dissertation). The American University.

Tallmadge, G. K. (1977). *The joint dissemination review panel idea book*. Washington, DC: National Institute of Education, and the US Office of Education.

Tausch, N. Hewstone, M., Schmid, K., Hughes, J., & Cairns, E. (2011). Extended contact effects as a function of closeness of relationship with ingroup contacts. *Group Processes & Intergroup Relations, 14*, 239–254.

Tausch, N., Hewstone, M., Kenworthy, J., Cairns, E., & Christ, O. (2007). Cross-community contact, perceived status differences, and intergroup attitudes in Northern Ireland: The mediating roles of individual-level vs. group-level threats and the moderating role of social identification. *Political Psychology, 28*, 53–61.

*Tavakoli, M., Hatami, J., & Thorngate, W. (2010). Changing stereotypes in Iran and Canada using computer mediated communication. *Journal of Intercultural Communication, 23*, 5–5.

Tropp, L. R., & Pettigrew, T. F. (2005). Relationships between intergroup contact and prejudice among minority and majority status groups. *Psychological Science, 16*, 651–653.

*Trubowitz, J. (1969). *Changing the racial attitudes of children.* New York, NY: Praeger.

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

U.S. Department of Justice (2013). *Hate crime statistics of the Federal Bureau of Investigation.* Retrieved from http://www.fbi.gov/about-us/cjis/ucr/hate-crime/2012

Viechtbauer, W**.** (2008). Analysis of moderator effects in meta-analysis. In J. Osborne (Ed.), *Best practices in quantitative methods* (pp. 471–487). Thousand Oaks, CA: Sage.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1–48.

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis*. Research Synthesis Methods, 1*, 112–125.

*Vittrup, B., & Holden, G. W. (2011). Exploring the impact of educational television and parent-child discussions on children's racial attitudes. *Analyses of Social Issues and Public Policy, 11*, 82–104.

Wagner, U., & Hewstone, M. (2012). Intergroup contact. In L. R. Tropp (Ed.), *The Oxford handbook of intergroup conflict* (pp. 193–209)**.** New York, NY: Oxford University Press.

Wagner, U., Christ, O., & Pettigrew, T. F. (2008). Prejudice and group related behaviour in Germany. *Journal of Social Issues, 64*, 403–416.

Wagner, U., Hewstone, M., & Machleit, U. (1989). Contact and prejudice between Germans and Turks: A correlational study. Human Relations, 42, 561–574.

*Warring, D., Johnson, D. W., Maruyama, G., & Johnson, R. T. (1985). Impact of

different types of cooperative learning on cross-ethnic and cross-sex relationships. *Journal of Educational Psychology, 77*, 53–59.

*Wayne, E. K. (2008). Is It Just Talk? Understanding and evaluating intergroup dialogue. *Conflict Resolution Quarterly, 25*, 451–478.

*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*, 597–608.

*Williams, D. H. (1935). *The effects of an interracial project upon the attitudes of Negro and White girls within the Y.W.C.A* (Unpublished master's thesis). Columbia University.

*Williams, E. (1972). *Effects of intergroup discussion on social distance and personal space of black and white students* (Unpublished doctoral dissertation). The University of Texas at Austin.

*Wilson, M. D. (2006). *Improving intergroup relations through programmed intergroup contact: A model process and its effects* (Unpublished doctoral dissertation). The Pennsylvania State University.

Wright, S. C. & Aron, A. (2010). The extended contact effect. In J. Levine, M. & A. Hogg (Eds.), *The encyclopedia of group processes and intergroup relations* (pp. 265–267). Thousand Oaks, CA: Sage.

Wright, S. C., Aron, A., McLaughlin-Volpe, T., & Ropp, S. A. (1997). The extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social Psychology, 73*, 73–90.

*Yablon, Y. B. (2010). Contact as a means of inter-religious engagement: The role of religious culture in peace-building activities. *International Handbooks of*

Religion and Education, 4, 497–512.

*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*, 249–263.

*Yablon, Y. B., & Katz, Y. J. (2001). Internet-based group relations: A high school peace education project in Israel. *Educational Media International*, *38*, 175–82.

Footnotes

[1] Within the context of this search component, we were thankfully assisted by Thomas F. Pettigrew (University of California, Santa Cruz) and Rupert Brown (University of Sussex, UK).

[2] None of the considered documents included a posttest that was conducted more than one year after the end of the intervention.

[3] When a dependent variable was reverse scored (i.e., a lower score indicates a better result), we changed the sign of the calculated effect size.

[4] We assume the absence of an interaction between the factors "intervention vs. control" and "time".

[5] We refrained from conducting a common multiple meta-regression test of the hypotheses since they are of an isolated nature and conceptually independent. In addition, a multiple meta-regression analysis could only be conducted with a reduced sample of comparisons. A listwise deletion would limit the number of the included cases since the tests of the different hypotheses are conceptually (see below) based on partially different samples of comparisons.

[6] The test offered by Begg and Mazumdar (1994) is based on the rank correlation between the standardized effect sizes and transformed standard errors. A significant rank correlation (Kendall's τ) signals an association between the two variables. In case of funnel plot asymmetry, high standard errors should be systematically associated higher effect sizes. The method provided by Egger et al. (1997) uses the inverse of the standard error to predict an index that is calculated as the effect size divided by the corresponding standard error. In case the intercept differs significantly from zero, results could be biased.

[7] Comparison 53 (externally standardized residual: 2.58), comparison 52 (externally standardized residual: 2.33), and comparison 79 (externally standardized residual: 2.23) in Supporting Information C, Table C1.

[8] Comparison 3 (externally standardized residual: 3.16) and comparison 40 (externally standardized residual: -2.10) in Supporting Information C, Table C2.

[9] Comparison 37 (externally standardized residual: 2.58) in Supporting Information C, Table C2.

[10] The *p*-value that corresponds with the Q_{model}-value was .0546. This test of the overall model is equivalent to the two-tailed test of the regression weight of the dichotomous moderator majority (0) vs. minority (1) that is conducted by use of the *z*-distribution ($z = -1.92$, two-tailed $p = .0546$). Since we have specified a one-sided alternative hypothesis, we halved the *p*-value so that it is equivalent to a one-tailed *z*-Test.

[11] Comparison 74 in Supporting Information C, Table C1.

Table 1

Description of the Included Comparisons

Variable and Values	Primary cluster (k = 79) k	%	Secondary cluster (k =50) k	%
Decade				
Before 1961	3	3.8	—	—
1961 – 1970	5	6.3	2	4.0
1971 – 1980	17	21.5	1	2.0
1981 – 1990	11	13.9	4	8.0
1991 – 2000	16	20.3	8	16.0
2001 – 2012	27	34.2	35	70.0
Type of document				
Published	42	53.2	26	52.0
Journal article	28	35.4	24	48.0
Book / book chapter	14	17.8	2	4.0
Unpublished	37	46.8	24	48.0
Dissertation / master thesis	35	44.3	17	34.0
Other unpublished	2	2.5	7	14.0
Country of the first author				
USA	50	63.3	19	38.0
Other	29	36.7	31	62.0
Type / subtype of intervention				
Direct	63	79.8	42	84.0
Contact meeting	39	49.4	42	84.0
Cooperative group learning	24	30.4	—	—
Indirect	16	22.3	6	12.0
Extended contact	8	10.1	—	—
Virtual contact	8	10.1	5	12.0
Direct and indirect	—	—	2	4.0
Contact meeting and virtual contact	—	—	2	4.0
Status				
Majority	39	49.4	14	28.0
Minority	20	25.3	19	38.0
Majority and minority	9	11.4	4	8.0
No status hierarchy	11	13.9	13	26.0
Context				
Protracted conflict	11	13.9	26	52.0
No protracted conflict	68	86.1	24	48.0

Table 1 (continued)

Description of the Included Comparisons

Variable and Values	Primary cluster (k = 79) k	%	Secondary cluster (k =50) k	%
Setting				
School	53	67.1	10	20.0
College / University	16	20.3	11	22.0
Other Setting	10	12.7	29	58.0
Duration of the intervention (gross time)				
1 – 7 Days	5	6.3	16	32.0
> 1 week – 1 month	12	15.2	18	36.0
> 1 month – 6 months	52	65.8	10	20.0
> 6 months – 12 months	4	5.1	6	12.0
Cannot be specified	6	7.6	—	—
Days with delivery of the intervention				
1 – 10 days	35	44.3	33	66.0
11 – 30 days	20	25.3	7	14.0
31 – 60 days	6	7.6	2	4.0
> 60 days	4	5.1	—	—
Cannot be specified	14	17.7	8	16.0
Net time (i.e., days with delivery multiplied by hours per day)				
1 – 10 hours	15	19.0	3	6.0
11 – 50 hours	37	46.8	15	30.0
51 – 100 hours	9	11.4	19	38.0
Net time (i.e., days of delivery multiplied with hours per day)				
> 100 hours	4	5.1	3	6.0
Cannot be specified	14	17.7	10	20.0
Country of implementation				
Austria	—	—	— 4	8.0
Australia	2	2.5	—	—
Canada	2	2.5	1	2.0
Cyprus	—	—	2	4.0
Germany	1	1.3	—	—
Great Britain	2	2.5	—	—
Israel	10	12.7	10	20.0
Malaysia	3	3.8	—	—
Netherlands	1	1.3	3	6.0
N. Ireland / Rep. of Ireland	4	5.1	—	—
Norway	2	2.5	—	—
Romania	1	1.3	—	—
South Africa	—	—	6	12.0
USA	45	57.0	21	
Mixed	6	7.6	3	6.0

Table 1 (continued)

Description of the Included Comparisons

Variable and Values	Primary cluster (k = 79) k	%	Secondary cluster (k =50) k	%
Age of the participants				
5 – 9 years	7	8.9	—	—
10 – 13 years	31	39.2	4	8.0
14 – 18 years	22	27.8	35	70.0
> 18 years	19	24.1	11	22.0
Sex of the participants (% female)				
0 – 30%	—	—	1	2.0
31 – 70%	48	60.8	35	70.0
71 – 100%	8	10.1	10	20.0
Cannot be specified	23	29.1	4	8.0
Dependent variable – level of generalization				
Known individuals	12	15.2	—	—
Target outgroup	38	48.1	31	62.0
Unspecified outgroup	16	20.3	15	30.0
Mixed	13	16.5	4	8.0
Dependent variable – content				
Cognitive	17	21.5	18	36.0
Affective/Behavioral	24	30.4	10	20.0
Mixed	38	48.1	22	44.0
Design				
Posttest only with control	12	15.2	—	—
Pretest-posttest with control	67	84.8	—	—
Pretest-posttest single group	—	—	50	100.0
Assignment to conditions				
Randomized (individuals)	19	24.1	—	—
Not randomized	60	75.9	—	—
No control group	—	—	50	100.0
Type of control				
No treatment	64	81.0	—	—
Placebo treatment	15	19.0	—	—
No control group	—	—	50	100.0
Type of posttests				
Only immediate (< 1 month)	71	89.9	33	66.0
Only delayed (1 – 12 months)	—	—	6	12.0
Direct and delayed	8	10.1	11	22.0
Interval between the end of the intervention and the immediate posttest				
1 – 7 days	60	75.9	37	84.1
> 1 week – less than 1 month	13	16.5	2	4.5
Cannot be specified	6	7.6	5	11.4

Table 1 (continued)

Description of the Included Comparisons

	Primary cluster (k = 79)		Secondary cluster (k =50)	
Variable and Values	k	%	k	%
Interval between the end of the intervention and the delayed posttest				
1 month – less than 6 smonths	4	50.0	12	70.6
6 months – 12 months	4	50.0	5	29.4
Total sample size of the comparison (immediate posttest)				
Up to 30	12	15.2	23	52.3
31 – 100	39	49.4	16	36.4
101 – 250	20	25.3	5	11.4
251 – 500	7	8.9	—	—
501 – 750	1	1.3	—	—
Mean attrition rate (immediate posttest)				
Up to 10 %	35	44.3	16	36.4
11 – 30%	21	26.6	11	25.0
31 – 50%	10	12.6	3	6.8
> 50 %	—	—	3	6.8
Cannot be specified	13	16.5	11	25.0
Attrition rate – intervention-control difference (immediate posttest)				
Up to 10 %	25	31.6	—	—
11 – 30%	3	3.8	—	—
31 – 50%	1	1.3	—	—
> 50 %		—	—	—
Attrition rate – intervention-control difference (immediate posttest)				
Cannot be specified	50	63.3	—	—
Total number of items				
1 – 10	17	21.5	20	40.0
11 – 50	45	57.0	19	38.0
51 – 100	12	15.2	9	18.0
> 100	1	1.3	—	—
Cannot be specified	4	5.1	2	4.0

Note. k = number of comparisons.

Table 2

General Effectiveness of Contact Interventions

Time of posttest	Cluster	$\hat{\mu}_\theta$	95% CI	*Q*	$\hat{\tau}^2$	I^2	*k*	*N*
Immediate	Primary	0.28***	[0.21, 0.35]	251.19***	0.06	70.72	79	9212
	Secondary	0.39***	[0.31, 0.47]	150.60***	0.05	73.96	44	2159
Delayed	Primary	0.23***	[0.13, 0.34]	9.11	0.00	10.42	8	1186
	Secondary	0.35***	[0.21, 0.50]	46.27***	0.05	61.84	17	464

Note. $\hat{\mu}_\theta$ = estimated average of the true effects; CI = confidence interval; *Q* = homogeneity statistic; $\hat{\tau}^2$ = estimated variance between the true effects; I^2 = amount of the true variance among the total variance; *k* = number of comparisons; *N* = total number of participants.

*** $p < .001$.

Table 3

Effectiveness of Contact Interventions as a Function of Type of Intervention, Subtype of Intervention, Context, Status, and Level of Generalization

	Primary cluster				Secondary cluster			
Variable	$\hat{\mu}_\theta$	95% CI	*k*	Q_{model}	$\hat{\mu}_\theta$	95% CI	*k*	Q_{model}
Type of intervention								
Direct	0.29***	[0.21, 0.37]	63	0.47	0.41***	[0.32, 0.50]	37	0.99
Indirect	0.23***	[0.07, 0.38]	16		0.33*	[0.05, 0.61]	5	
Direct and indirect	—	—	—		0.24	[–0.12, 0.60]	2	
Subtype of intervention								
Meeting	0.31***	[0.21, 0.41]	39	5.82	0.41***	[0.32, 0.50]	37	0.99
Cooperative	0.25***	[0.13, 0.38]	24		—	—	—	
Extended	0.42***	[0.19, 0.65]	8		—	—	—	
Virtual	0.08	[–0.13, 0.28]	8		0.33*	[0.05, 0.61]	5	
Meeting and virtual	—	—	—		0.24	[–0.12, 0.60]	2	
Context								
Protracted conflict	0.31***	[0.14, 0.48]	11	0.15	0.47***	[0.35, 0.60]	20	3.75
No protracted conflict	0.27***	[0.19, 0.35]	68		0.34***	[0.22, 0.46]	24	
Status								
Majority	0.38***	[0.27, 0.49]	39	4.39*[10]	0.46***	[0.29, 0.63]	12	0.53
Minority	0.20**	[0.05, 0.35]	20		0.37***	[0.22, 0.53]	15	
Level of Generalization								
Individual	0.26**	[0.08, 0.44]	12	0.51	—	—	—	1.67
Target outgroup	0.32***	[0.22, 0.43]	38		0.45***	[0.34, 0.55]	25	
Unspecified outgroup	0.21**	[0.05, 0.38]	16		0.33***	[0.19, 0.47]	15	

Note. $\hat{\mu}_\theta$ = estimated average of the true effects; CI = confidence interval; *k* = number of comparisons; Q_{model} = test whether the average true effects differ between the levels of the moderator.

* $p < .05$. ** $p < .01$. *** $p < .001$.

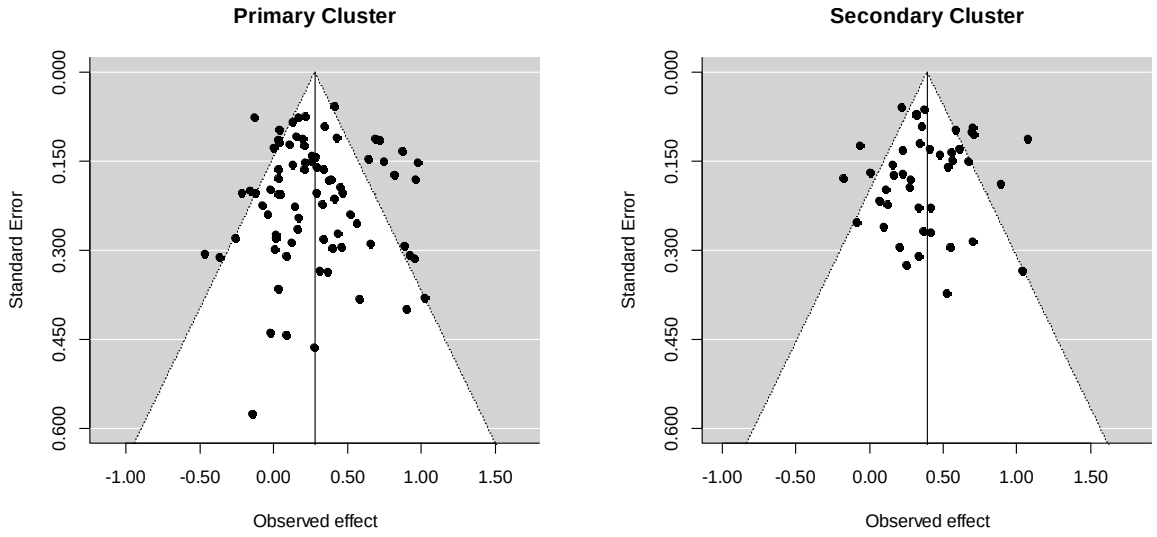

Figure 1. Funnel plots for the immediate posttest in the primary ($k = 79$) and secondary ($k = 44$) cluster. The points represent the included intervention-control comparisons. They display the observed effect size on the abscissae and the corresponding standard error on the ordinate. The estimated average true effect is indicated by a vertical line