Sequential models of intergroup contact and social categorization:

An experimental field test of integrated models

Loris Vezzali[a], Elena Trifiletti[b], Ralf Wölfer[c], Gian Antonio Di Bernardo[d], Sofia Stathi[e],

Veronica Margherita Cocco[d], Alessia Cadamuro[d], Soraya Elizabeth Shamloo[b], Miles

Hewstone[f]

[a] Faculty of Medicine, University of Modena and Reggio Emilia, viale Allegri 9, 42121,

Reggio Emilia, Italy

[b] Department of Human Sciences, University of Verona, Via San Francesco 22,

37129, Verona, Italy

[c] Department of Consensus & Conflict, DeZIM Institute, Mauerstraße 76, 10117 Berlin,

Germany

[d] Department of Human Sciences and Education, University of Modena and Reggio Emilia,

viale Allegri 9, 42121, Reggio Emilia, Italy

[e] Department of Psychology, Social Work and Counselling, University of Greenwich, Old

Royal Naval College, London SE10 9LS, UK

[f] Department of Experimental Psychology, Oxford University, Woodstock Road, Oxford,

OX2 6GG, UK.

Correspondence concerning this article should be addressed to Loris Vezzali, Faculty of

Medicine, University of Modena and Reggio Emilia, viale Allegri 9, 42121, Reggio Emilia,

Italy. Email: loris.vezzali@unimore.it.

Sequential Models of Intergroup Contact and Social Categorization:

An Experimental Field Test of Integrated Models

Abstract

Research has proposed different models of how the contact situation should be structured to maximize contact effects, focusing in particular on the role of categorization during contact. We conducted two experimental field interventions (*Ns* = 247 and 247) to test models that integrate different levels of categorization. Each of the tested models was contrasted against a no-intervention control condition. In both studies, we assessed effects shortly after the intervention (one week later, i.e., posttest) and then after approximately six months (i.e., follow-up). In the first study, results generally support the model where categorization precedes decategorization, showing effects on major dependent measures highlighted in research on intergroup contact: quantity and quality of contact, cross-group friendships, intergroup anxiety (marginal effect at follow-up), outgroup attitudes (only at follow-up). Evidence for follow-up effects for this model was, however, weaker in Study 2, where the delayed effects of the intervention emerged only indirectly, via changes in contact quality, outgroup attitudes, and approach behavioral intentions at posttest. Comparisons of the other two models (decategorization then categorization; and simultaneous categorization and decategorization) with the control condition (only in Study 1) provided weaker and inconsistent results.

Keywords: intergroup contact, decategorization, categorization, prejudice reduction, intergroup relations.

Impressive evidence supports Allport's (1954) 'contact hypothesis', whereby intergroup contact can reduce prejudice, and this effect is even stronger when optimal contact conditions (i.e., equal status, cooperation, pursuit of common goals, institutional support) are met (Hodson & Hewstone, 2013; Pettigrew & Tropp, 2006; Vezzali & Stathi, 2017). One important omission from Allport's (1954) list of optimal conditions, albeit an omission that has only come to light decades later, is that it does not specify how contact should be structured in terms of social categorization in order to maximize its effects. Allport was, however, clearly aware of the process of 'subtyping', whereby individuals exposed to outgroup members who challenge their stereotypes may simply treat those exemplars as atypical, place them in a separate subtype of outgroup members, and resist changing their stereotype. In order to address this limitation, scholars have proposed different, seemingly incompatible models emphasizing personalization or decategorization (interaction as individuals, focused on personal qualities and characteristics; Brewer & Miller, 1984), categorization (keeping categories salient during contact; Hewstone & Brown, 1986), or recategorization (merging groups into a single superordinate group; Gaertner, Mann, Murrell, & Dovidio, 1989) during contact. However, others have argued that these models are not necessarily incompatible, and can be ordered in a longitudinal sequence which, if fulfilled, would allow the maximization of contact effects (Hewstone, 1996; Pettigrew (1998). R. Brown and Hewstone (2005) proposed a further integrative model, arguing that interpersonal and group differences should be stressed simultaneously. In addition, another approach deriving from R. Brown and Hewstone's (2005) model seems plausible, one in which categorization is followed by decategorization.

In this article, we present the first experimental test of these integrative models. We focus on categorization and decategorization, rather than recategorization. The competing predictions between R. Brown and Hewstone's (2005) and Pettigrew's (1998; see also Hewstone, 1996) models do not refer to recategorization, and there is little doubt that where recategorization (or preferably 'dual identity', in which valued subordinate identities are retained, but in a framework of a shared superordinate identity; Gaertner & Dovidio, 2000) is introduced, it should be in the final stage of successful contact. To test hypotheses, we conducted two experimental field interventions designed to improve relations between majority (Italian) and minority (immigrant) members in an educational context. In addition, departing from most research in the field and responding to the need for studies with delayed outcome measures (see Paluck, Green, & Green, 2019), we tested effects both immediately (1 week) and later in time (six months) after the interventions ended.

Models of contact and categorization

We focus on two models that have attempted to explain the optimal level of categorization within the contact situation. These models draw largely on the interpersonal-intergroup continuum posited by social identity theory (SIT; Tajfel & Turner, 1979). According to Tajfel and Turner (1979), when individuals encounter members of different groups and are operating at the interpersonal end of the continuum, only individual differences are salient, thus their interaction is driven exclusively by individual characteristics. When, instead, individuals encounter members of different groups while operating at the intergroup end of the continuum, group differences are salient and the interaction is entirely determined by differences characterizing their respective groups.

According to the decategorization model (Brewer and Miller, 1984, 1988), since prejudice stems from social categorization (Tajfel & Turner, 1979), individuals from different groups should interact as single individuals (interpersonal end of the interpersonal-intergroup continuum). Interpersonal interactions allow individuals to "attend to information that replaces category identity as the most useful basis for classifying each other" (Brewer & Miller, 1984, p. 288). In this way, prejudice is no longer relevant, as individuals' interactions are only determined by their personal qualities and characteristics. This model received support from multiple laboratory experiments (e.g., Bettencourt, Brewer, Croak, & Miller, 1992; Miller, Brewer, & Edwards, 1985) and from research on cross-group friendships (Davies, Tropp, Aron, Pettigrew, & Wright, 2011). However, the model suffers two main conceptual limitations. First, individuals may be unwilling or unable to relinquish their identities (Hewstone, 1996). Second, if decategorization is successful, the generalization of positive attitudes to the outgroup as a whole is impeded, because the outgroup member is not cognitively associated with their larger outgroup category.

According to the intergroup contact model (originally referred to as the 'mutual intergroup differentiation' model) proposed by Hewstone and Brown (1986; R. Brown & Hewstone, 2005), group salience should be maintained during contact, so that individuals can associate the individual outgroup member with their larger outgroup category, and therefore generalize the positive attitudes developed during contact to the outgroup as a whole. This model, in turn, received empirical support from experimental (e.g., R. Brown, Vivian, & Hewstone, 1999, Study 1; van Oudenhoven, Groenewoud, & Hewstone, 1996), correlational (e.g., Harwood, Hewstone, Paolini, & Voci, 2005; Voci & Hewstone, 2003),

and longitudinal (e.g., Binder et al., 2009; Greenland & Brown, 1999, Study 2) research. However, this model itself has two main limitations. First, if contact is negative, negative outgroup attitudes may be generalized to the group as a whole (Graf & Paolini, 2017; Islam & Hewstone, 1993). Second, group salience during contact may provoke anxiety, which is one of the main barriers to positive relations between groups (Stephan, 2014).

Integrative Contact Models

Sequential Models of Contact and Categorization

As noted above, some scholars have proposed that rather than choosing between these seemingly incompatible models, they could be successfully combined, to maximize contact effects. Hewstone (1996) argued that, especially when intergroup relations are divisive, individuals should interact first on an interpersonal basis, before introducing typicality and membership salience to ensure that contact is sufficiently intergroup to yield generalized effects. In conflictual contexts, such a temporal strategy should avoid the danger of fomenting intergroup divisions.

Pettigrew (1998) similarly elaborated the idea of combining approaches, and suggested that the different levels of categorization could be ordered in a longitudinal sequence. In the initial phase of contact (interpersonal level, decategorization), individuals should relate to one another on the basis of their personal characteristics, in order to maintain low levels of intergroup anxiety (which might disrupt the newly initiated contact). In the second phase (group level, categorization), when the risk of arousing intergroup anxiety is lower, group salience should be introduced, in order to foster generalization of outgroup attitudes to the outgroup as a whole. In the third phase, Pettigrew (1998) argued

6

that recategorization should be introduced, although he noted that this could not always be achieved.

Eller, Abrams, and colleagues (Eller & Abrams, 2003, 2004; Eller, Abrams, & Koschate, 2017) tested Pettigrew's (1998) integrative model with correlational and longitudinal designs. In their first study (Eller & Abrams, 2003), conducted with U.S. students studying Spanish in Mexico and considering relations between U.S. Americans and Mexicans, the authors found that contact at time 1 (one week after their arrival in Mexico) was characterized more strongly by interpersonal and dual identity than by categorization and recategorization. At time 2 (one week later), the interpersonal level was the highest, whereas the difference among the other three levels was nonsignificant. These results were replicated in two longitudinal studies (Eller & Abrams, 2004, Studies 1 and 2), considering the relationship between English and French people (Study 1) and Mexicans and U.S. Americans (Study 2), respectively.

The results of these three studies are not consistent with Pettigrew's integrative model, since, according to that model, the interpersonal level should decrease, and the intergroup level should increase over time, whereas the recategorization level should be higher in the final stage. However, in the first study (Eller & Abrams, 2003), it was difficult to define when the first contact phase ended and the second phase started, and Pettigrew (1998) highlighted that contact phases may overlap. Moreover, the latter two studies (Eller & Abrams, 2004, Studies 1 and 2) did not specifically assess the initial stage of contact, and therefore do not qualify as stringent tests of the theory. But most importantly, Pettigrew did not argue that contact naturally follows the proposed longitudinal sequence; rather, he

7

proposed that contact effects will be stronger *when* the proposed longitudinal sequence is followed.

Eller et al. (2017) conducted a longitudinal study with three waves (separated by six-month intervals), investigating relations between high-school students in Germany. Specifically, the authors measured contact and prejudice between parallel school classes within each grade in each tested school. The study focused on three levels of categorization: intergroup, superordinate identity, and dual identity. The authors tested whether the effectiveness of the levels of categorization depends on the external structure provided by the social context. Specifically, levels of categorization should be maximally effective when they match the external structure. In the context examined, time 1 was characterized by segregation between school classes within a school (students always attended classes with the same set of classmates); in contrast, at times 2 and 3 the structure was of integration between classes (classes were not fixed, but within each grade students could, depending on their chosen courses, attend classes with peers that had belonged to different school classes in the past). Results supported Eller et al.'s hypotheses, revealing that the intergroup level was associated with reduced intergroup anxiety and increased desire for contact when the external structure was one of segregation (between time 1 and time 2; in this case, students belonged to distinct school classes). In contrast, when the external structure changed (from time 1 to time 2) and was one of integration (between time 2 and time 3), the dual identity level was associated with reduced ingroup bias. Although not specifically designed to test Pettigrew's (1998) integrated model, this study provides important indications, specifically regarding the fact that the levels of categorization may differentially predict outgroup attitudes depending on the contact phase. To provide a

comprehensive test of Pettigrew's model, however, it is necessary to manipulate levels of categorization experimentally and test *when* contact is most effective in reducing prejudice.

R. Brown and Hewstone's integrative contact model

R. Brown and Hewstone (2005) proposed a second integrative model involving a combination of their intergroup contact model (R. Brown & Hewstone, 2005; Hewstone & Brown, 1986) and the decategorization model (Brewer & Miller, 1984, 1988). In particular, in line with the notion that interpersonal and intergroup dimensions may be orthogonal (Stephenson, 1981), the integrative model proposed that group salience is not incompatible with interactions at the interpersonal level. In other words, according to this model, contact should be maximally effective when both interpersonal and intergroup characteristics are salient during contact. In fact, while the interaction at the interpersonal level should lower anxiety stemming from interacting with an outgroup member, group salience should favor generalization of contact effects to the outgroup as a whole.

Experimental evidence consistent with this model was gleaned from two laboratory studies conducted by Ensari and Miller (2002). The first study, using a sample of Turkish non-religious university students, manipulated self-disclosure and typicality of the outgroup member (a confederate presented as an Islamic student). Results revealed that outgroup attitudes were more positive when both self-disclosure and outgroup typicality were high, and, therefore, when the interpersonal and intergroup levels of categorization were simultaneously salient. Results were replicated and extended in the second study, which investigated the relationship between liberals and conservatives within a sample of U.S. American students. The study showed that outgroup attitudes were more positive when

high self-disclosure was paired with high typicality of the outgroup member or with high group salience.

Additional indirect evidence is provided by correlational studies showing that the effects of cross-group friendships (which imply high levels of self-disclosure, and therefore an interaction with the outgroup member at the interpersonal level; Davies et al., 2011) on prejudice reduction are stronger when group salience/outgroup typicality are high (Poteat, 2015; Vonofakou, Hewstone, & Voci, 2007). However, in a correlational study considering non-disabled workers' attitudes toward disabled colleagues, aimed at testing whether contact effects were moderated by interpersonal and/or intergroup levels of categorization, there were no significant three-way interactions between contact and the interpersonal and intergroup levels on outcome variables (intergroup anxiety and empathy, explicit and implicit outgroup attitudes; Vezzali, 2008).

The supportive evidence provided for this model has limitations. First, the experimental studies by Ensari and Miller (2002) that provide support for this model lack ecological validity, since it is not clear whether the two levels of categorization (interpersonal, intergroup) can be simultaneously activated and/or maintained over time, in a newly initiated contact situation, and when important group identities are at play. Second, evidence from correlational studies is not only scarce but also mixed.

A novel integrative sequential model of intergroup contact

Our review of the contact models presented above led us to propose a new integrative sequential model in which categorization precedes decategorization. This model can be extrapolated from the core premise of Hewstone and Brown's (1986; R. Brown & Hewstone, 2005) intergroup contact model, namely that categorization is unavoidable and

10

represents a key ingredient of successful contact (in terms of generalization of contact effects). As argued by Hewstone and Brown, individuals may be unable or unwilling to relinquish important social identities. In addition, attitude generalization is unlikely without group salience. Therefore, introducing categorization from the beginning of a contact experience can be in line with individuals' motivation as well as being more likely to make contact effective, in the sense of achieving generalized attitude change. A potential risk of decategorized contact is, as noted, that any positive effects of contact may not generalize. A possible solution may be to introduce categorization at the beginning of the contact experience in order to increase the chances of generalization; once categorization has been introduced, contact can be decategorized in order to mitigate the risk that categorized contact may increase intergroup anxiety (Islam & Hewstone, 1993; Stephan, 2014). This sequence should reduce the likelihood of contact participants using negative outgroup stereotypes and attitudes as the basis for intergroup evaluations. This new model, where categorization precedes decategorization, implies that categorization continues to remain salient over time, while decategorization allows a "relaxation" of the intergroup relationship, leading to attitude change.

Initial evidence for this hypothesis was provided by Van Oudenhoven et al. (1996). These authors asked Dutch students to work cooperatively with a Turkish confederate and systematically varied when group salience was introduced by referring to the ethnicity of the confederate: at the beginning and again during a break between the tasks (first experimental condition), or only during the break between tasks (second experimental condition); in the no-intervention control condition, ethnicity was not mentioned at all. Results revealed that attitude generalization (evaluation of Turkish people as a whole) was

not different in the two experimental conditions, but in both cases was higher than in the

control condition. Therefore, once individuals are categorized as group members, they can

continue being associated with their larger outgroup category, even if in later phases contact

also becomes characterized by interpersonal characteristics. Van Oudenhoven et al. (1996)

did not manipulate decategorization of contact in later phases of their study; hence, in the

present research, in Study 1 we build on their research by explicitly manipulating

decategorization after a phase of categorization. This primacy of the categorization phase is

likely to make it difficult, if not impossible, to subtype outgroup members who have the

potential to change the outgroup stereotype (Hewstone, 1994).

The Present Research

We designed two experimental intervention studies aimed at testing Pettigrew's

(1998; see also Hewstone, 1996) and R. Brown and Hewstone's (2005) integrative contact

models in the field, compared to a no-intervention control condition. In addition, we also

tested the effect of a third intervention, a novel integrative sequential contact model where

categorization precedes decategorization.

To test the models, we conducted two field experiments, in which we tested the

effects of three categorization-based interventions and their persistence over time. In both

studies, we assessed the impact of experimentally manipulated contact on the relationship

between majority (Italian) and minority (immigrant) group members, considering majority

participants only (due to the size of the available minority sample). In addition, to provide a

stringent and comprehensive test of Pettigrew's (1998) model, we investigated the impact

of initial experiences of contact on respondents who had thus far had little or no contact

with their new classmates, by investigating first-year high-school students enrolled in

mixed classes (composed of both majority and minority members) who had just started attending the high school.

Given the relative absence of direct empirical evidence for the effectiveness of integrative contact models, we decided to test each model against a control condition. In the present case, it should be noted that the control group is not a no-intervention control group. In fact, all participants have the same opportunity for contact, but in the control group there is no manipulation of level of categorization. This represents a strict test of our hypothesis, since all participants engaged to a greater or less extent in contact, and the only aspect that varied was the manipulated level of categorization. Therefore, any effects of each condition compared with the no-intervention control condition are likely to be due to the experimental manipulation, rather than to different levels of intergroup contact.

In the first study, we created three experimental conditions that systematically varied interpersonal and intergroup levels of categorization, by asking participants to engage in tasks aimed at reinforcing the manipulation. The first condition ('decategorization-categorization') followed Pettigrew's (1998) integrative model, by presenting, first, decategorization (interpersonal level) and then categorization (intergroup level). The second and third conditions created two different models, both of which can be derived from R. Brown and Hewstone's (2005) model. The second condition ('categorization-decategorization') presented, first, categorization and then decategorization; the third condition ('categorization+decategorization') presented categorization and decategorization simultaneously. In a fourth, control condition, participants received no instructions on how they should relate with their classmates.

This experimental design enabled us to test predictions derived from both Pettigrew's (1998) sequential model (first experimental condition, decategorization-categorization) and two models derived from Hewstone and Brown's (1986; R. Brown & Hewstone, 2005) intergroup contact model.

In the first study, students completed a questionnaire before the experimental manipulation (pretest), one week after the end of the manipulation (posttest), and at the end of the school year (follow-up; approximately six months later). Although long-term assessments are relatively rare in contact research (cf. Lemmer & Wagner, 2015), delayed posttests are of primary importance. As pointed out by Paluck et al. (2019), policy makers need to know about the persistence of contact effects, that is, whether effects last days, weeks, months, or even years. In their meta-analysis, Paluck et al. were only able to locate 27 studies that, amongst other requirements, assessed outcomes at least one day after the contact intervention began (and only 17 studies could be included if the requirement was to assess outcomes at least one day after the intervention *ended*).

As dependent variables in post-intervention assessments, we included key variables investigated by contact research, being various measures of what a contact-based intervention should be seeking to increase (quantity and quality of contact, cross-group friendships, outgroup attitudes, and behavioral intentions) or decrease (intergroup anxiety).

We also conducted a follow-up study, based on indications provided by the first study. Since in Study 1 the second experimental condition (categorization-decategorization), representing the novel integrative contact model proposed in this article, was the most successful, in Study 2 we sought to replicate this result by conducting a similar experimental field intervention among first-year high-school majority and minority

14

students, in which we compared categorization-decategorization with a no-categorization, no intervention control condition. Study 2 also differed from Study 1 in two other ways. First, we did not administer a pretest questionnaire. Second, we added a measure of behavioral intentions.

Study 1

Methods

Sample and Design.

Sample size was determined by: a) the number of schools that authorized participation in the study, and b) the number of students who agreed to participate (having also obtained consent from their parents/guardians). Participants were students from 18 classes of five different schools in northern Italy, enrolled in the first year of high school. All five schools were diverse in terms of students' national backgrounds and the percentage of minority students varied between approximately 30% and 40%. We included only majority group students (i.e., Italians) as participants, because (a) most minority group students had insufficient linguistic ability to complete the administered questionnaire independently and (b) the resulting sample size of minority group students (i.e., non-Italians) was too small to be considered for statistical analysis. In the present three-wave study, a total of $N = 247$ students participated at the pretest (mean age = 14.22, $SD = 0.77$; 118 males, 113 females, 16 missing data), $N = 236$ at the posttest (96%; mean age = 14.22, $SD = 0.77$; 113 males, 107 females, 13 missing data), and $N = 222$ at the follow-up (90%; mean age = 14.20, $SD = 0.76$; 108 males, 100 females, 14 missing data). Data at all three measurement points was available from $N = 213$ students (86%; mean age = 14.20, $SD = 0.77$; 101 males, 98 females, 14 missing data).

Participants ($N = 247$) were randomly allocated to one of three experimental conditions (see Design and procedure section): decategorization followed by categorization (decategorization-categorization, $n = 55$); categorization followed by decategorization (categorization-decategorization, $n = 54$); categorization and decategorization administered simultaneously (categorization+decategorization, $n = 72$); and a no-intervention control condition ($n = 66$). Allocation to each condition was made at the class level (classes, rather than individual students, were randomly assigned to each condition).

Ethics.

Ethical approval for the research was provided by the Ethics Committee at [blinded for review.

Procedure.

The researchers who conducted the intervention were students enrolled in educational academic courses at a northern Italian university. All researchers were trained by the first author of the present article.

Between approximately one and two weeks after the beginning of the school year, participants completed a questionnaire (pretest). Immediately after filling it in, we administered the first part of the experimental manipulation.

In the decategorization-categorization condition, we provided participants with the first set of instructions on how to interact with their new classmates in the following two weeks by focusing on interpersonal characteristics (instructions for the three intervention groups are provided in Supplemental Materials). After explaining instructions, we gave participants 36 questions from the closeness generation procedure (Aron, Melinat, Aron, Vallone, & Bator, 1997). They were told that they did not necessarily need to use these

16

questions reciprocally to interact with classmates; rather, the questions could be used to better understand how to relate with peers at the interpersonal level. With the assistance of the researchers, participants also engaged in various activities and games aimed at reinforcing the manipulation. For instance, they were invited to respond to reciprocal personal questions in pairs in front of the class. In the "Guess who" game, they were asked to write on a sheet three personal characteristics that defined them, and if they wished to add a self-portrait. Then, in turn they would extract one sheet and read the three characteristics, while the rest of the class should guess to whom the sheet referred. One week after the first set of instructions, in order to reinforce the manipulation, researchers came back to each class to continue activities aimed at decategorizing the contact. Two weeks after the first set of instructions, to further reinforce the manipulation, participants were asked to write an essay on what they had learned about their classmates in the two previous weeks.

After writing the essay, we gave participants the second set of instructions, aimed at categorizing the contact. Participants were asked to engage in activities similar to those performed in the two previous weeks, this time, however, focusing on intergroup rather than on interpersonal characteristics. For instance, in a "Guess where" game, participants were asked to write on a sheet three characteristics of their ethnic group; in this case, the class had to guess which country each sheet referred to. One week after the first set of instructions, researchers conducted another session engaging participants in similar activities; two weeks after the second set of instructions, participants were asked to write an essay on what they had learned about their classmates in the two previous weeks. All sessions lasted approximately one hour.

The procedure for the categorization-decategorization condition was similar. In this case, however, the order of decategorization and categorization instructions was reversed.

In the categorization+decategorization condition, we merged categorization and decategorization instructions, and we repeated them in both the first and second set of instructions. Games matched the double nature of these instructions. For instance, in the "Guess who and where" game, participants were asked to write on a sheet both three personal characteristics and three characteristics of their ethnic group, and the class had to guess who the person was, and their group.

Participants in the control condition were simply asked, in the same timeframe as the three intervention conditions, to write the two essays on what they had learned about their classmates in the two previous weeks.

Participants from all groups completed a questionnaire during class time approximately one week (posttest) and six months (follow-up) after the last intervention session.

Measures.

Data were collected during the school year 2015/2016. At each phase of the intervention, students answered questions that assessed relevant socio-demographic items as well as the measures of intergroup relations described below.

Contact quantity.

Participants reported the frequency of outgroup contact on six items (e.g., "How much contact do you have with immigrants?"; e.g., Islam & Hewstone, 1993; see also Lolliot et al., 2015) using a five-point scale (1 = *none or never* to 5 = *very much or always*).

The mean of these six items represented participants' contact quantity at each respective phase ($\alpha_{pretest}$ = .81, $\alpha_{posttest}$ = .84, $\alpha_{follow-up}$ = .83).

Contact quality.

Participants reported the quality of outgroup contact on four bipolar scales ("When you come into contact with immigrants, how do you find that contact?": "hostile/friendly," "indifferent/caring," "competitive/cooperative," "rude/kind"; see Capozza, Trifiletti, Vezzali, & Favara, 2013; Vezzali et al., 2018) using a five-point scale. Higher scores indicate higher quality of contact at each respective phase ($\alpha_{pretest}$ = .84, $\alpha_{posttest}$ = .85, $\alpha_{followup}$ = .83) scales.

Cross-group friendships.

Participants reported their number of outgroup friends on five items (e.g., "How many immigrant friends do you have?"; e.g., R. N. Turner, Hewstone, & Voci, 2007; see also Lolliot et al., 2015) using a five-point scale (1 = *none*, 2 = *1-2*, 3 = *3-4*, 4 = *5-6*, 5 = *more than six*). The mean of these five items represented participants' cross-group friendships at each respective phase ($\alpha_{pretest}$ = .70, $\alpha_{posttest}$ = .68, $\alpha_{follow-up}$ = .73) scales.

Intergroup anxiety.

Participants reported their level of intergroup anxiety on twelve items, related to how they feel when they think about interaction with immigrants (e.g., "worried;" six items were reverse coded, e.g., "confident"; see Stephan & Stephan, 1985) using a five-point scale (1 = *not at all* to 5 = *very much*). The mean of these twelve items represented students' intergroup anxiety at each respective phase ($\alpha_{pretest}$ = .86, $\alpha_{posttest}$ = .85, $\alpha_{follow-up}$ = .82) scales.

Outgroup attitudes.

We used the General Evaluation Scale (Wright, Aron, McLaughlin-Volpe, & Ropp, 1997) to measure attitudes. Participants reported their level of intergroup attitudes on six bipolar items (e.g., "negative/positive," "suspicious/trusting") using a five-point scale. Higher scores indicated more positive outgroup attitudes. The mean of these six items represented participants' intergroup attitudes at each respective phase ($\alpha_{pretest}$ = .88, $\alpha_{posttest}$ = .89, $\alpha_{follow-up}$ = .87) scales.

Statistical analyses.

In preliminary analyses, we compared participants whose responses over time could ('matched') versus could not ('unmatched') be matched over time, to test whether participants who dropped out after pretest and those who dropped out after posttest were significantly different from those having data at the three collection phases. To this end, we applied independent *t*-tests on sociodemographic characteristics (age, gender) and on pretest outcome variables.

Given the clustered nature of our data (participants were students drawn from 18 different classes) and the fact that random assignment to the intervention conditions occurred at the class level, further preliminary analyses checked for the nested structure of data by computing intra-class correlations (ICCs) and class-level variance of each dependent variable at the pretest, posttest, and follow-up.

Next, we conducted multilevel regression analyses in Mplus (8.2; Muthén & Muthén, 2017), in order to assess the effects of the intervention on dependent variables at posttest and follow-up, while controlling for class-level variance (i.e., level 2 variance). We conducted the same regression analyses on pretest variables to check whether random assignment to conditions was successful. In order to test the effect of each experimental

20

condition against the control condition, the four conditions were coded using dummy coding with three dummy variables; in each dummy variable, the control condition (coded as 0) was contrasted with one of the three intervention conditions (coded as 1). To control for class-level variance, we used the cluster option in Mplus.

Results and Discussion

Initial data inspection.

Dropout analyses comparing matched and unmatched respondents revealed that respondents with matched data did not differ from either participants who dropped out after the pretest (posttest dropout) or participants who dropped out after the posttest (follow-up dropout), ts ≤ 1.31, ns, in terms of age ($d_{posttest\ dropout}$ = 0.09; $d_{follow-up\ dropout}$ = 0.25), contact quantity ($d_{posttest\ dropout}$ = 0.21; $d_{follow-up\ dropout}$ = 0.06), contact quality ($d_{posttest\ dropout}$ = 0.10; $d_{follow-up\ dropout}$ = 0.03), cross-group friendship ($d_{posttest\ dropout}$ = 0.21; $d_{follow-up\ dropout}$ = 0.02), intergroup anxiety ($d_{posttest\ dropout}$ = 0.23; $d_{follow-up\ dropout}$ = 0.28), outgroup attitudes ($d_{posttest\ dropout}$ = 0.22; $d_{follow-up\ dropout}$ = 0.12), or gender, χ^2s(1) < 1, ns, ($\varphi_{posttest\ dropout}$ = 0.02; $\varphi_{follow-up\ dropout}$ = .04).

Descriptive statistics at pretest, posttest, and follow-up across experimental conditions are reported in Table 1. Table 2 shows ICCs and class-level variance scores for dependent variables at pretest, posttest, and follow-up. For some of the variables, ICCs were > .05 and class-level variance was significant or approached significance; both results suggested that the clustered structure of our data could affect the results. Therefore we analyzed the data using multilevel regression models.

Intervention effects.

Results of multilevel regression analyses are summarized in Table 3. At the pretest, there were no systematic differences between the control condition and any intervention condition, thus confirming successful random assignment.

At posttest, the only significant difference between the decategorization-categorization and the control condition concerned intergroup anxiety, which was significantly lower in the decategorization-categorization condition compared to the control. There was also a marginally significant difference in relation to contact quantity and outgroup attitudes, both of which were slightly higher in the decategorization-categorization condition compared to the control. No significant difference was observed between the categorization+decategorization condition and the control condition. In contrast, significant differences between the categorization-decategorization condition and the control condition emerged on all the dependent variables except outgroup attitudes. Contact quantity, contact quality, and cross-group friendship were higher (and intergroup anxiety was lower) in the intervention condition compared to the control.

At follow-up, we found no significant differences between the control condition and either the decategorization-categorization or the categorization+decategorization condition, except a marginal tendency for cross-group friendships to be higher in the decategorization-categorization than in the control condition. Significant differences between the categorization-decategorization condition and the control condition emerged on contact quantity, contact quality, cross-group friendship, and outgroup attitudes, which were all higher in the intervention condition compared to the control. For intergroup anxiety, the effect approached conventional levels of significance, with the mean slightly lower in the intervention compared to the control condition.[1]

In sum, Study 1 compared the effects of three different categorization-based interventions against a control group both at the end of the intervention and after six months. In doing so, we demonstrated that categorization-decategorization (where categorization was introduced before decategorization) was the only condition to be consistently effective.

Study 2

The results of Study 1 showed that the condition in which categorization preceded decategorization was the only condition that yielded reliable and consistent improvements, compared with the control condition. We therefore continued to focus on this particular intervention strategy in Study 2. Study 2 pursued two main research objectives: first, given the renewed attention to replication in our discipline, we sought to replicate the intervention effects from Study 1 using a new sample; and second, we examined specific time differences in more detail with the help of a larger intervention group.

To maintain comparability with Study 1, we again focused on the beginning phase of a new contact situation, and specifically with newly formed first-year high-school classes, considering the relationship between Italians and immigrants from the point of view of Italian participants. Participants were unlikely to have had regular contact with their new classmates (although they were might have had intergroup contact in the past). In this study, we did not administer a pretest questionnaire, in order to comply with school requests to simplify the procedure and limit disruption and time lost to teaching by our data collection. Moreover, Study 1 had revealed that our procedures led to successful randomization of participants to conditions. However, we included the two posttest questionnaires, one week after the intervention and approximately six months later, in order

to replicate findings on the persistence of effects found in Study 1 and provide robust evidence for the efficacy of the tested model. Departing from Study 1, which had assessed any contact with immigrants, in this study we assessed contact in school specifically, to test the effects of the intervention on actual relations with classmates. We also included a measure of behavioral intentions, which are stronger predictors of behavior (Fishbein & Ajzen, 2010), and especially of cross-group behavior (Vezzali, Stathi, Giovannini, Capozza, & Visintin, 2015), than are attitudes.

Methods

    Sample and Design.

    Sample size was determined by: a) the number of schools that authorized participation in the study, and b) the number of students who agreed to participate (having also obtained consent from their parents/guardians). Participants ($N = 247$) were students from 15 classes of four different schools in Italy (different classes from these schools also participated in Study 1), enrolled in the first year of high school. Three participants were excluded because, despite both their parents were Italian, they declared to be immigrants (see Procedure section). As in Study 1, we included only majority group students (i.e., Italians) as participants for the same reasons outlined above. In the present study, we assessed the dependent measures immediately after the intervention (posttest) and after six months (follow-up). The posttest questionnaire was completed by $N = 201$ participants (82%; mean age $= 14.17$, $SD = 0.64$, 124 males, 77 females). At follow-up, data were available from $N = 164$ participants (6/%; mean age $= 14.12$, $SD = 0.73$, 98 males, 66 females). Participants were randomly allocated to one of two conditions, either the categorization-decategorization condition or the control condition. Allocation to each

condition was made at the class level (classes as a whole were randomly assigned to each condition). Both conditions were tested in each school.

Ethics.

Ethical approval for the research was provided by the Ethics Committee at [blinded for review.

Procedure.

The procedure was identical to that used in Study 1, with one main difference: in this study, we did not administer a pretest questionnaire. Students were asked to complete a questionnaire one week after the last intervention session (posttest questionnaire) and approximately six months later (follow-up). At posttest, participants had to self-select whether they belonged to the Italian group, by selecting the appropriate link.

Measures.

Data were collected in a school-based online survey, conducted in the respective schools within the regular school setting, with the help of trained administrators during the school year 2016/2017. At both posttest and follow-up, students filled in a questionnaire that assessed relevant sociodemographic variables, as well as the measures of intergroup relations described below. Unless noted otherwise, measures were identical to those used in Study 1.

Contact quantity.

Similar to Study 1, but with items asking specifically about contact at school, participants reported the frequency of contact on six items (e.g., "How much contact do you have with immigrants at school?"; see Lolliot et al., 2015) using a five-point scale (1 =

none or never to 5 = *very much or always*). The mean of these six items represented

participants' contact quantity at posttest and follow-up ($\alpha_{posttest}$ = .80, $\alpha_{follow-up}$ = .85).

Contact quality.

Participants reported the quality of outgroup contact at school on four items

(common stem: "When you come into contact with immigrants from your school, how do

you find that contact?"; item-specific completion: "hostile/friendly," "indifferent/caring,"

"competitive/cooperative," "rude/kind"; Capozza et al., 2013) using a five-point scale;

higher scores indicated greater quality of contact. The mean of these four items represented

participants' contact quality at posttest and follow-up ($\alpha_{posttest}$ = .77, $\alpha_{follow-up}$ = .83).

Cross-group friendships.

As in Study 1, but asking specifically about friendships at school, participants

reported their number of outgroup friends on three items (e.g., "How many immigrant

friends do you have at school?"; see Lolliot et al., 2015; Turner et al., 2007) using a five-

point scale (1 = *none*, 2 = *1-2*, 3 = *3-4*, 4 = *5-6*, 5 = *more than six*). The mean of these three

items represented participants' cross-group friendships at posttest and follow-up ($\alpha_{posttest}$

= .78, $\alpha_{follow-up}$ = .81) scales.

Intergroup anxiety.

Participants reported their level of intergroup anxiety when they think about

interacting with immigrants on twelve items, using a five-point scale (1 = *not at all* to 5 =

very much). The mean of these twelve items represented students' intergroup anxiety at

posttest and follow-up ($\alpha_{posttest}$ = .85, $\alpha_{follow-up}$ = .89) scales.

Outgroup attitudes.

Participants reported their outgroup attitudes toward immigrants on six bipolar items (e.g.: "negative/positive" "suspicious/trusting") developed by Wright et al. (1997), using a five-point scale. Higher scores indicated more positive outgroup attitudes. The mean of these six items represented participants' outgroup attitudes at posttest and follow-up (Cronbach's α_{posttest} = .87, Cronbach's $\alpha_{\text{follow-up}}$ = .91) scales.

Behavioral intentions.

In addition to the measures included in Study 1, participants reported hypothetical behavioral intentions to immigrants on nine items (common stem: "In general, when thinking of immigrants, I want to…") using a six-point scale (0 = *not at all* to 5 = *extremely*), which formed three subscales including three items each: aggressive behavioral intentions (item-specific completion, e.g., "confront them"), approach behavioral intentions (item-specific completion, e.g., "find out more about them"), and avoidance behavioral intentions (item-specific completion, e.g., have nothing to do with them"; see Mackie, Devos, & Smith, 2000; Tam, Hewstone, Harwood, Voci, & Kenworthy, 2006). However, while approach and avoidance behavioral intentions had satisfactory reliabilities (approach: Cronbach's α_{posttest} = .88, Cronbach's $\alpha_{\text{follow-up}}$ = .83; avoid: Cronbach's α_{posttest} = .80, Cronbach's $\alpha_{\text{follow-up}}$ = .88), the scale for aggressive behavioral intentions showed inadequate internal consistency (Cronbach's α_{posttest} = .37, Cronbach's $\alpha_{\text{follow-up}}$ = .47) and was, therefore, excluded from all following analyses.

Statistical Analyses.

We inspected differences between matched and unmatched (i.e., those who did and did not complete measures at follow-up, respectively) participants on posttest dependent variables and socio-demographic characteristics.

We tested for effects of the intervention with multilevel regression analyses in Mplus (8.2; Muthén & Muthén, 2017). As in Study 1, the control condition was coded as 0 and the intervention condition as 1. Afterwards, we explored possible delayed effects of the intervention by testing a path analysis regression model with observed variables in Mplus. In more detail, we examined the effect of the experimental condition on posttest scores of contact, attitudes, and approach intentions that, in turn, were related to follow-up scores of the same constructs, while controlling for the autoregressive effects and covariances. The model included seven observed variables: experimental condition, and contact quality, outgroup attitudes, approach intentions at posttest and follow-up. We tested indirect effects of the intervention on dependent variables at follow-up, via contact quality, outgroup attitudes, and approach intentions at posttest, with Bias Corrected and accelerated (BCa) 95% bootstrapping confidence intervals (5,000 resamples).

Results and Discussion

Initial data inspection.

Dropout analyses comparing 'matched' (with data at both posttest and follow-up) versus 'unmatched' (with data only at posttest) participants revealed small differences on age, $t(197) = 2.07$, $p = .04$, $d = 0.37$, and contact quantity, $t(199) = 2.03$, $p = .04$, $d = 0.40$, while differences on the other variables were nonsignificant, ts $(199) \leq 1.83$, ps $\geq .07$ (contact quality: $d = 0.01$; cross-group friendship: $d = 0.33$; intergroup anxiety: $d = 0.30$; outgroup attitudes: $d = 0.12$; approach: $d = 0.01$: avoidance: $d = 0.18$).

Descriptive statistics at posttest and follow-up in the experimental and control conditions are reported in Table 4. Table 5 shows ICCs and class-level variance scores for outcome variables at posttest and follow-up. As in Study 1, for some of the variables ICCs

were > .05 and class-level variance was significant or approached significance, suggesting that the clustered structure of the data should be taken into account in further analyses.

Intervention effects.

Results of multilevel modeling are reported in Table 6. At posttest, findings revealed a significant difference between the intervention and control group on three dependent variables, and an effect approaching significance on a fourth dependent variable. Participants in the intervention group reported higher contact quality, more positive outgroup attitudes, and higher approach behavioral intentions. We found an effect approaching significance for avoidance behavioral intentions, with participants in the intervention group reporting slightly lower avoidance intentions than control participants. No significant differences emerged for contact quantity, cross-group friendship, and intergroup anxiety.

At follow-up, no significant difference emerged between the intervention and the control group, suggesting that the intervention was successful in the short term, with relatively powerful effects at posttest which, however, weakened over time. We therefore explored possible delayed intervention-based effects. That is, while the main effect of condition was no longer significant at follow-up, we examined whether the evidential posttest effects indirectly accumulated into meaningful changes at follow-up. To assess this, we examined the effect of experimental condition on participants' posttest scores and, in turn, the effect of these posttest scores on participants' follow-up scores, while controlling for autoregressive effects and covariances. To this end, we tested a path analysis focusing on the three variables that were significantly affected by the intervention at posttest (contact quality, outgroup attitudes, and approach behavioral intentions).

29

Longitudinal path analysis.

To test delayed effects of the intervention, we estimated a path analysis with observed variables, that examined the effect of the experimental condition (coded 1 for experimental condition and 0 for control condition) on posttest contact quality, outgroup attitudes, and approach intentions. We estimated auto-regressive and cross-lagged paths from posttest variables to follow-up variables. The fit of this model was satisfactory: $\chi^2(3)$ = 3.36, p = .34; CFI = 1.00; RMSEA = .02; SRMR = .02. Figure 1 summarizes the model.

As already shown by the multilevel regression, experimental condition had a significant positive effect on posttest contact quality and outgroup attitudes, as well as a marginal effect on approach intentions. Moreover, both contact quality and outgroup attitudes at posttest were significantly associated with improved outgroup attitudes and contact quality at follow-up, respectively. Finally, contact quality at posttest significantly predicted approach intentions at follow-up.

Indirect effects tested with BCa 95% confidence intervals are summarized in Table 7. Contact quality at posttest significantly mediated the effects of the intervention on all three dependent variables at follow-up. Indirect effects via outgroup attitudes at posttest were significant in relation to contact quality and outgroup attitudes at follow-up. Finally, approach intentions at posttest only mediated the effect of the intervention on approach intentions at follow-up.

In sum, Study 2, like Study 1, revealed an improvement in intergroup relations, evidenced by positive effects of the categorization intervention (specifically, categorization followed by decategorization) compared to a control group using a posttest – follow-up

design; importantly, effects of the intervention at follow-up emerged only indirectly, via improvements obtained at posttest.

General Discussion

We conducted two experimental field interventions to test competing predictions made by various models that have proposed how contact and categorization should be combined most effectively (R. Brown & Hewstone, 2005; Hewstone, 1996; Pettigrew, 1998). We evaluated both short- and longer-term consequences of the interventions, considering multiple dependent variables central to contact research (quantity and quality of contact, cross-group friendships, intergroup anxiety, outgroup attitudes, and both approach and avoidance behavioral intentions). In Study 1, we tested three different, theoretically-derived integrative models of contact against a no-intervention control condition. Results provided stronger support for the model in which individuals engage first in categorization and then in decategorization (categorization-decategorization) than for two alternative models. Evidence was weaker in Study 2, where we contrasted only the most effective categorization-decategorization condition against a control condition. Although some effects emerged at posttest, but not follow-up (approximately six months later), we did detect indirect effects at follow-up, via variables that emerged as significant at posttest. Although the effects detected were not especially strong, a quite common occurrence in field research (Anderson, Lindsay, & Bushman, 1999), made more likely by the use of demanding follow-up tests, the research provides important evidence of the value of addressing social categorization in interventions designed to promote intergroup contact. We, first, summarize and discuss the main findings, considering their importance in terms of both theoretical models of social categorization and contact, and their importance to

practitioners and policy makers. Next, we consider how to maximize the effects of contact, and close by acknowledging some limitations of these studies, and highlighting future research directions.

We believe that effects obtained on the different types of contact (quantity, quality, cross-group friendships) are of particular relevance. They also add to increasing literature investigating predictors of contact (Stathi et al., 2020). Whereas outgroup attitudes represent the key outcome in most research on intergroup contact, since contact, and especially high-quality contact, is a powerful predictor of prejudice reduction, finding that an intervention fosters more frequent and positive cross-group interactions represents direct evidence of the creation of more positive intergroup relations, an outcome that is likely to produce long-term effects which are reinforced over time. Effects obtained on behavioral intentions (in Study 2) are similarly encouraging, given that intentions are more powerful predictors of future behavior than are attitudes (Fishbein & Ajzen, 2010). As noted by Paluck et al. (2019; see also Ülger, Dette-Hagenmeyer, Reichle, & Gaertner, 2018), these tests are of primary importance to practitioners and policy makers, who, in order to consider contact as a viable and effective strategy for integration policies, need to know that effects persist over time, therefore contributing effectively to sustained social integration.

The results that emerged for the key categorization-decategorization condition do not, however, allow overly optimistic conclusions. In Study 1, effects at posttest were nonsignificant for outgroup attitudes; and at follow-up, effects were only marginal for intergroup anxiety. In Study 2, effects at posttest were marginal for avoidance intentions, and nonsignificant for contact quantity, cross-group friendships, and intergroup anxiety.

32

Most importantly, direct effects of the manipulation were no longer significant at follow-up. Nonetheless, indirect effects (on contact quality, outgroup attitudes, and approach intentions) did emerge, indicating that future studies should better clarify both the *consistency* and *duration* of contact effects, which can be done by focusing on how to maximize the effects of contact. The inconsistency of effects on specific variables between the two studies may be due to a number of factors, considering that the studies were conducted at a distance of one year from one another. However, we believe that the main result of interest is that in both studies the same categorization manipulation had effects on key outcomes identified by contact research.

Maximizing contact effects

Theorists who developed models of contact and categorization sought to maximize contact effects. They debated primarily the roles of categorization and decategorization. The novel integrative model, in which categorization precedes decategorization seems, according to data obtained in our two studies, to hold promise as a way to combine the relative advantages of the two strategies and at the same time minimize their disadvantages. Fostering categorization at the beginning of contact ensures that ingroup members do associate outgroup members with their broader category and, once this is done, do not lose sight of social categories during the intergroup interaction.

Surprisingly, we found only weak evidence for the causal sequence in which decategorization preceded categorization (proposed by both Hewstone, 1996, and Pettigrew, 1998). There may be various explanations for this finding. For example, locating contact at the interpersonal level from the outset may have severed the link between known outgroup individuals and the whole outgroup; thereby, introducing group salience in later

phases may not have had the desired effect, since known outgroup members had already been perceived in terms of their individual characteristics, and perhaps already been subtyped. Both Hewstone (1996) and Pettigrew (1998) argue that decategorizing contact may serve to reduce anxiety that can characterize the beginning of intergroup interactions. However, intergroup anxiety was not high in the present studies (cf. Tables 1 and 2); therefore, participants may not have had any 'need' to reduce anxiety, and decategorizing contact may have acted as on obstacle to generalization. Future studies specifically focused on this model should test whether decreases in intergroup anxiety can produce changes in outgroup attitudes over time. Decategorizing initial contact may be more important, even necessary, in highly conflictual settings, where initial anxiety is high (cf. Hewstone, 1996). In such contexts, reinforcing group differences, even if this is done within a positive climate, may be detrimental to subsequent relations. This possibility is consistent with the temporal integrated model of intergroup contact and threat (Abrams & Eller, 2017), which states that contact effects are shaped by the specific situation and its extant threats. It should also be noted that our operationalization might not fully capture Pettigrew's (1998) longitudinal model. One way to interpret the proposed sequence of that model is that interpersonal processes should be replaced over time by intergroup processes (this is what we did). An alternative interpretation would be adding intergroup processes over time, so that the second stage is characterized by both a focus on intergroup characteristics *and* dyadic contact based on interpersonal characteristics. In the latter case, our operationalization would not fully capture Pettigrew's model.

We also did not find evidence for the integrated contact model proposed by R. Brown and Hewstone (2005), which proposed making both interpersonal and intergroup

characteristics simultaneously salient. It is possible that individuals are unable, or find it difficult, to consider simultaneously interpersonal and intergroup characteristics. This possibility is consistent with self-categorization theory (J. C. Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). According to the functional antagonism principle, when group differences are salient, interpersonal characteristics are inhibited, and vice versa. Note, however, that, although in this condition participants were instructed to consider both group memberships and personal characteristics simultaneously from the outset of the contact, we cannot rule out that both interpersonal and intergroup dimensions were simultaneously salient in other conditions. For instance, in the condition in which categorization precedes decategorization, after having framed known outgroup members in terms of their group characteristics in the first phase, participants may have 'added' a focus on interpersonal characteristics during this phase; since there was time for the categorization phase to 'sink in', the additional focus on interpersonal characteristics may not have had disruptive effects on participants' judgments. This is also consistent with Pettigrew's (1998) contention that stages may overlap.

We stated explicitly at the outset that our studies were *not* aimed at comparing between integrative models. In the absence of experimental evidence for their effectiveness, we sought, rather, to test whether each model was effective when compared against a no-categorization information control condition. The control condition, in which contact could occur to the same extent as in the other conditions, provided a baseline of contact effects; our manipulation allowed us to test the added benefit of a manipulation aimed at altering categorization over time which could, and, as our findings show, in fact did affect contact. Certain combinations of categorization and decategorization affect the amount and type of

35

contact that people engage in and how they experience it. Therefore, results do not allow conclusions on which model is the "best." Rather, they enable us to demonstrate that the model in which categorization precedes decategorization yielded effects across two studies (especially, when evaluated one week after the end of the intervention), compared with the control.

Our conclusions on the importance of raising the salience of intergroup differences before interpersonal characteristics should also be qualified by the nature of our experimental approach. The fact that we did not have a "pure" categorization condition does not allow us to conclude that it is important to implement decategorization after categorization: possibly, simply implementing categorization at the beginning of the intervention is sufficient. Future studies should compare a categorization-decategorization condition with a categorization-only condition, to understand the exact role played by decategorization processes.

We believe the use of a no intervention-control condition increases confidence in our findings. The control condition exemplifies what generally occurs in mixed classes, where contact is not explicitly structured. Our studies provide evidence of the benefits of structuring contact based on social categorization. It should be noted that this represents a conservative approach, since non-structured contact may have produced some effects per se; with a no-contact control condition.

Most research on intergroup contact has been correlational (Pettigrew & Tropp, 2006), although there has been a recent marked rise in longitudinal studies (e.g., Binder et al., 2009; Christ et al., 2014; Swart et al., 2011; see also Christ & Wagner, 2013). In addition, evidence for contact effects has also been provided by experimental research

conducted in the lab (e.g., Gaertner et al., 1989; Van Oudenhoven et al., 1996). Much less

frequent has been theoretically-driven research in the field based on face-to-face contact

(Paluck & Green, 2009; but see Cameron & Turner, 2017; for an example of an

experimental intervention using indirect, or not face-to-face contact, see Tercan et al.,

2021), or research on theoretically structured experimental interventions that has included

long-term assessment of outcome variables (Paluck et al., 2019). The interventions

investigated in this article specifically address this question in a dynamic way, by

examining how theoretically-driven activities can be structured over time to maximize the

effects of contact, therefore bridging theory and practice in the field of prejudice reduction,

and increasing confidence in the power of contact to improve intergroup relations (R. N.

Turner & Cameron, 2016). These studies did, however, have some limitations, which

should be acknowledged and may guide future research.

Limitations and future directions

Some of the limitations we should acknowledge are specifically tied to the decision

to conduct experimental field interventions, rather than more highly-controlled lab studies.

Although we believe our manipulations generally reflect the theoretical models we sought

to test, they represent, of course, only some of the possible operationalizations of

constructs. Laboratory studies might be deemed more suitable than field interventions for

experimental tests of specific hypotheses. A particular risk in field studies is a loss of

control, especially in terms of adherence of participants to the manipulation, leading to

lower internal validity. As an example, although we provided instructions and activities

(needed to clarify instructions and also to motivate participants in this context), we could

not control whether participants adhered to instructions in the absence of researchers.

Other limitations are more specific, and we highlight five.

First, although participants were invited to focus on interpersonal and/or intergroup characteristics (depending on phase and experimental condition) during school interactions, we have no control over how participants behaved in, and whether they carried over these instructions to, out-of-school interactions. Related to this point, it is possible that instructions in the categorization+decategorization condition conveyed more extensive information, causing a cognitive overload that may have stymied their efficacy. Note, however, that we conducted activities with participants to ensure their understanding of the task; any cognitive overload may, however, relate to the difficulty of simultaneously considering both interpersonal and intergroup aspects. Second, although we also included measures of behavioral intentions in Study 2, the dependent variables were all self-reports. Future studies should also include behavioral observations complementing the self-reported results. Third, due to practical constraints, random allocation to the intervention conditions was done at the class, rather than the participant, level. Although we controlled for class-level variance with multilevel modeling, a full randomization approach would have maximized control against confounding variables and non-independence of data (it would, however, have risked contamination of the independent variable). Fourth, we did not check for school-level variance in our analyses given the small number of schools involved. Fifth, it can be argued that assessing contact with respect to classmates would have provided more direct evidence for the effect of the manipulation. We deliberately chose, however, to begin (Study 1) with general contact measures, since we aimed to investigate the effects of the intervention on intergroup relations more generally (as it is practice in contact

research); however, we focused on contact with classmates in Study 2, to also test effects at the level of the specific setting.

These limitations could, to some extent, have been mitigated in laboratory experiments. However, the experimental field studies we report have the valuable commodity of ecological validity. In the real world, intergroup contact may need time to develop, and artificial contexts may struggle to provide the significance, depth and time needed for manipulations to take effect, and for the different stages of contact to develop. In addition, if intergroup contact is to be used as a strategy that has demonstrable impact in real-world settings, such as schools, then experimental field studies such as the ones we conducted provide crucial evidence for policy makers.

To conclude, decategorization and categorization strategies are not mutually exclusive. Rather, they can be complementary, and contact may be most effective when they are combined in an integrated sequential contact model that conceives them as different (and, potentially, partly overlapping) stages of intergroup contact.

Endnotes

1. To account for pretest differences between participants, we also tested multilevel
   random coefficient regression models including pretest scores as within (first-level)
   predictors (both for posttest and follow-up dependent variables). Results generally
   replicated findings of multilevel regression models reported in Table 3, except for
   follow-up contact quality, for which the difference between control and
   categorization-decategorization conditions was nonsignificant ($p = .37$). Moreover,
   the difference between this intervention condition and the control condition was
   marginally significant for contact quality at posttest ($p = .052$) and for cross-group
   friendship at follow-up ($p = .08$). Results are available upon request to the first
   author.

Declaration of conflicting interest

The authors declare that there are no conflicts of interest.

Data sharing

The data for this work are available on Open Science Framework at:

https://osf.io/as2j7/?view_only=70fcfbb30fd54c8e93c4b0cab3899174, doi:

10.17605/OSF.IO/AS2J7

References

Abrams, D., & Eller, A. (2017). A temporally integrated model of intergroup contact and threat (TIMICAT). In L. Vezzali & S. Stathi (Eds.), *Intergroup contact theory: Recent developments and future directions* (pp. 72-91). Abingdon, UK: Routledge.

Allport, G. W. (1954). *The nature of prejudice*. New York, NY: Addison-Wesley.

Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science, 8*, 3-9.

Aron, A., Melinat, E., Aron, E., Vallone, R. D., & Bator, R. J. (1997). The experimental generation of interpersonal closeness: A procedure and some preliminary findings. *Personality and Social Psychology Bulletin*, *23*, 363-377. doi:10.1177/0146167297234003

Bettencourt, B. A., Brewer, M. B., Croak, M. R., & Miller, N. (1992). Cooperation and the reduction of intergroup bias: The role of reward structure and social orientation. *Journal of Experimental Social Psychology*, *28*, 301-319. doi:10.1016/0022-1031(92)90048-O

Binder, J., Zagefka, H., Brown, R., Funke, F., Kessler, T., & Mummendey, A. (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. doi:10.1037/a0013470

Brewer, M. B., & Miller, N. (1984). Beyond the contact hypothesis: Theoretical perspectives on desegregation. In N. Miller & M. B. Brewer (Eds.), *Groups in contact: The psychology of desegregation* (pp. 281-302). New York, NY: Academic Press.

Brewer, M. B., & Miller, N. (1988). Contact and cooperation: When do they work? In P.

Katz & D. Taylor (Eds.), *Eliminating racism: Means and controversies* (pp. 315-326).

New York, NY: Plenum.

Brown, R., & Hewstone, M. (2005). An integrative theory of intergroup contact. *Advances*

in Experimental Social Psychology, *37*, 255-343. doi:10.1016/S0065-2601(05)37005-5

Brown, R., Vivian, J., & Hewstone, M. (1999). Changing attitudes through intergroup

contact: The effects of group membership salience. *European Journal of Social*

Psychology, *29*, 741-764. doi:10.1002/(SICI)1099-0992(199908/09)29:5/6<741::AID-

EJSP972>3.0.CO;2-8

Brown, T. A. (2006). *Confirmatory factor analysis*. New York, NY: Guilford.

Cameron, L., & Turner, R. N. (2017). Intergroup contact among children. In L. Vezzali &

S. Stathi (Eds.), *Intergroup contact theory: Recent developments and future directions*

(pp. 151-168). Abingdon, UK: Routledge.

Capozza, D., Trifiletti, E., Vezzali, L., & Favara, I. (2013). Can contact improve humanity

attributions? *International Journal of Psychology*, *48*, 527-541.

doi:10.1080/00207594.2012.688132

Christ, O., Schmid, K., Lolliot, S., Swart, H., Stolle, D., Tausch, N….& Hewstone, M.

(2014). Contextual effect of positive intergroup contact on outgroup prejudice.

Proceedings of the National Academy of Sciences, *111*, 3996-4000.

doi:10.1073/pnas.1320901111

Christ, O., & Wagner, U. (2013). Methodological issues in the study of intergroup contact:

Towards a new wave of research. In G. Hodson & M. Hewstone (Eds.), *Advances in*

intergroup contact (pp. 233-261). New York, NY: Psychology Press.

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. doi:10.1177/1088868311411103

Eller, A., & Abrams, D. (2003). "Gringos" in Mexico: Cross-sectional and longitudinal effects of language school-promoted contact on intergroup bias. *Group Processes and Intergroup Relations*, *6*, 55-75. doi:10.1177/1368430203006001012

Eller, A., & Abrams, D. (2004). Come together: Longitudinal comparisons of Pettigrew's reformulated intergroup contact model and the common ingroup identity model in Anglo-French and Mexican-American contexts. *European Journal of Social Psychology*, *34*, 229-256. doi:10.1002/ejsp.194

Eller, A., Abrams, D., & Koschate, M. (2017). Can stateways change folkways? Longitudinal tests of the interactive effects of intergroup contact and categorization on prejudice. *Journal of Experimental Social Psychology*, *72*, 21-31. doi:10.1016/j.jesp.2017.04.005

Ensari, N., & Miller, N. (2002). The out-group must not be so bad after all: The effect of disclosure, typicality, and salience on intergroup bias. *Journal of Personality and Social Psychology*, *83*, 313-329. doi:10.1037/0022-3514.83.2.313

Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. New York, NY: Psychology Press.

Gaertner, S. L., & Dovidio, J. F. (2000). *Reducing intergroup bias: The common ingroup identity model*. Philadelphia, PA: Psychology Press.

Gaertner, S., Mann, J., Murrell, A., & Dovidio, J. F. (1989). Reducing intergroup bias: The benefits of recategorization. *Journal of Personality and Social Psychology*, *57*, 239-249. doi:10.1037/0022-3514.57.2.239

Graf, S., & Paolini, S. (2017). Investigating positive and negative intergroup contact. In L. Vezzali & S. Stathi (Eds.), *Intergroup contact theory: Recent developments and future directions* (pp. 92-113). Abingdon, UK: Routledge.

Greenland, K., & Brown, R. (1999). Categorization and intergroup anxiety in contact between British and Japanese nationals. *European Journal of Social Psychology*, *29*, 503-521.doi:10.1002/(SICI)1099-0992(199906)29:4<503::AID-EJSP941>3.0.CO;2-Y

Harwood, J., Hewstone, M., Paolini, S., & Voci, A. (2005). Grandparent-grandchild contact and attitudes toward older adults: Moderator and mediator effects. *Personality and Social Psychology Bulletin*, *31*, 393-406. doi:10.1177/0146167204271577

Hewstone, M. (1994). Revision and change of stereotypic beliefs: In search of the elusive subtyping model. *European Review of Social Psychology*, *5*, 69-109. doi:10.1080/14792779543000020

Hewstone, M. (1996). Contact and categorization: Social psychological interventions to change intergroup relations. In C. N. Macrae, C. Stangor, & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 323-368). New York, NY: Guilford Press. [Reprinted in C. Stangor (Ed.) (2000), *Stereotypes and prejudice.* Philadelphia, PA: Psychology Press].

Hewstone, M., & Brown, R. (1986). Contact is not enough: An intergroup perspective on the contact hypothesis. In M. Hewstone & R. Brown (Eds.), *Contact and conflict in intergroup encounters* (pp. 1-44). Oxford, UK: Blackwell.

Hodson, G., & Hewstone, M. (Eds.) (2013). *Advances in intergroup contact*. New York, NY: Psychology Press.

Hu, L., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55. doi:10.1080/10705519909540118

Islam, M. R. & Hewstone, M. (1993). Dimensions of contact as predictors of intergroup anxiety, perceived out-group variability, and out-group attitude: An integrative model. *Personality and Social Psychology Bulletin*, *19*, 700-710. doi:10.1177/0146167293196005

Jöreskog, K. G., & Sörbom, D. (2006). *LISREL 8.8 for Windows* [*Computer software*]. Skokie, IL: Scientific Software International, Inc.

Lemmer, G., & Wagner, U. (2015). Can we really reduce ethnic prejudice outside the lab? A meta-analysis of direct and indirect contact interventions. *European Journal of Social Psychology*, *45*, 152-168. doi:10.1002/ejsp.2079

Lolliot, S., Fell, B., Schmid, K., Woelfer, R., Swart, H., Voci, A., Christ, O…& Hewstone, M. (2015). Measures of intergroup contact. In G. Boyle, D. H. Saklofske, & G. Matthews (Eds.), *Measures of personality and social psychological constructs* (pp. 652-683). San Diego, CA: Academic Press.

Mackie, D. M., Devos, T., & Smith, E. R. (2000). Intergroup emotions: Explaining offensive behavioural tendencies in an intergroup context. *Journal of Personality and Social Psychology*, *79*, 602-616. doi:10.1037/0022-3514.79.4.602

Miller, N., Brewer, M. B., & Edwards, K. (1985). Cooperative interaction in desegregated settings: A laboratory analogue. *Journal of Social Issues*, *41*, 63-79. doi:10.1111/j.1540-4560.1985.tb01129.x

Muthén L, & Muthén B. (2017). *Mplus user's guide, 8th ed*. Los Angeles, CA: Muthén & Muthén.

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. doi:10.1146/annurev.psych.60.110707.163607

Paluck, E. L., Green, S. A., & Green, D. P. (2018). The contact hypothesis re-evaluated. *Behavioural Public Policy*, *3*, 129-158. doi:10.1017/bpp.2018.25

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, *49*, 65-85. doi:10.1146/annurev.psych.49.1.65

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*, 751-783. doi:10.1037/0022-3514.90.5.751

Poteat, V. P. (2015). Individual psychological factors and complex interpersonal conditions that predict LGBT-affirming behavior. *Journal of Youth and Adolescence*, *44*, 1494-1507. doi:10.1007/s10964-015-0257-5

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1-36. doi:10.18637/jss.v048.i02

Satorra, A., & P. M. Bentler (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507-514. doi:10.1007/BF02296192

Stathi, S., Pavetich, M. E., Di Bernardo, G. A., Cadamuro, A., Cocco, V. M., & Vezzali, L. (2020). Intergroup fears and concerns among minority and majority groups: Implications for contact and attitudes. *Journal of Community Psychology*, *48*, 1010-1027. doi:10.1002/jcop.22322

Stephan, W. G. (2014). Intergroup anxiety: Theory, research, and practice. *Personality and Social Psychology Review*, *18*, 239-255. doi:10.1177/1088868314530518

Stephan, W. G., & Stephan, C. W. (1985). Intergroup anxiety. *Journal of Social Issues*, *41*, 157-175. doi:10.1111/j.1540-4560.1985.tb01134.x

Stephenson, G. M. (1981). Intergroup bargaining and negotiation. In J. C. Turner & H. Giles (Eds.), *Intergroup behaviour* (pp. 168-198). Oxford, UK: Blackwell.

Swart, H., Hewstone, M., Christ, O., & Voci, A. (2011). Affective mediators of intergroup contact: A three-wave longitudinal study in South Africa. *Journal of Personality and Social Psychology*, *101*, 1221-1238. doi:10.1037/a0024450

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-47). Monterey, CA: Brooks/Cole.

Tam, T., Hewstone, M., Harwood, J., Voci, A., & Kenworthy, J. (2006). Intergroup contact and grandparent-grandchild communication: The effects of self-disclosure on implicit and explicit biases against older people. *Group Processes and Intergroup relations*, *9*, 413-429. doi:10.1177/1368430206064642

Tercan, M., Bisagno, E., Cocco, V. M., Kaçmaz, T., Turnuklu, A., & Vezzali, L. (2021). Reducing prejudice toward Syrian refugee children: A vicarious contact intervention

among Turkish elementary school children. *Journal of Community Psychology*, *49*, 564-587. doi:10.1002/jcop.22480

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, UK: Blackwell.

Turner, R. N., & Cameron, L. (2016). Confidence in contact: A new perspective on promoting cross-group friendship among children and adolescents. *Social Issues and Policy Review, 10*, 212-246. doi:10.1111/sipr.12023

Turner, R. N., Hewstone, M., & Voci, A. (2007). Reducing explicit and implicit outgroup prejudice via direct and extended contact: The mediating role of self-disclosure and intergroup anxiety. *Journal of Personality and Social Psychology*, *93*, 369-388. doi:10.1037/0022-3514.93.3.369

Ülger, Z., Dette-Hagenmeyer, D. E., Reichle, B., & Gaertner, S. L. (2018). Improving outgroup attitudes in schools: A meta-analytic review. *Journal of School Psychology*, *67*, 88-103. doi:10.1016/j.jsp.2017.10.002

van Oudenhoven, J. P., Groenewoud, T., & Hewstone, M. (1996). Cooperation, ethnic salience and generalization of interethnic attitudes. *European Journal of Social Psychology*, *26*, 649-661. doi:10.1002/(SICI)1099-0992(199607)26:4<649::AID-EJSP780>3.0.CO;2-T

Vezzali, L. (2008). *Theories of social identities: Experimental studies*. Doctoral dissertation thesis, Padova, Italy.

Vezzali, L., & Stathi, S. (Eds.). (2017). *Intergroup contact theory: Recent developments and future directions*. Abingdon, UK: Routledge.

Vezzali, L., Stathi, S., Giovannini, D., Capozza, D., & Visintin, E. P. (2015). "And the best essay is…": Extended contact and cross-group friendships at school. *British Journal of Social Psychology*, *54*, 601-615. doi:10.1111/bjso.12110

Vezzali, L., Turner, R., Capozza, D., & Trifiletti, E. (2018). Does intergroup contact predict personality? A longitudinal study on the bidirectional relationship between intergroup contact and personality traits. *European Journal of Social Psychology*, 48, 159-173. doi:10.1002/ejsp.2420010202

Voci, A., & Hewstone, M. (2003). Intergroup contact and prejudice toward immigrants in Italy: The mediational role of anxiety and the moderational role of group salience. *Group Processes and Intergroup Relations*, *6*, 37-54. doi:10.1177/1368430203006001011

Vonofakou, C., Hewstone, M., & Voci, A. (2007). Contact with outgroup friends as a predictor of meta-attitudinal strength and accessibility of attitudes toward gay men. *Journal of Personality and Social Psychology*, *92*, 804-820. doi:10.1037/0022-3514.92.5.804

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. doi:10.1037/0022-3514.73.1.73

Table 1. Descriptives of study variables at pretest, posttest, and follow-up across experimental conditions (Study 1).

Outcome variable	Decategorization-categorization			Categorization-decategorization			Categorization+ decategorization			Control group		
	M	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
Contact Quantity												
Pretest	2.66	0.74	55	2.81	0.87	54	2.59	0.80	71	2.53	0.82	63
Posttest	2.77	0.74	55	2.95	0.87	53	2.60	0.77	67	2.56	0.80	61
Follow-up	2.58	0.65	46	2.86	0.84	52	2.36	0.79	61	2.38	0.74	62
Contact Quality												
Pretest	3.77	0.83	55	3.84	0.82	52	3.69	0.85	71	3.67	0.92	63
Posttest	3.61	0.84	54	3.84	0.84	53	3.63	0.89	67	3.59	0.77	61
Follow-up	3.58	0.85	46	3.81	0.80	52	3.47	0.93	61	3.52	0.74	62
Cross-group Friendships												
Pretest	2.87	0.83	55	2.90	0.89	54	2.79	0.76	71	2.82	0.81	63
Posttest	2.96	0.70	55	3.18	0.85	53	2.90	0.78	67	2.81	0.83	61
Follow-up	2.87	0.66	46	3.06	0.79	53	2.58	0.82	61	2.57	0.82	62
Intergroup Anxiety												
Pretest	2.12	0.64	55	2.11	0.68	54	2.15	0.75	71	2.15	0.73	63
Posttest	2.13	0.73	54	2.16	0.64	53	2.34	0.70	67	2.46	0.75	61
Follow-up	2.25	0.61	46	2.18	0.69	53	2.56	0.73	61	2.38	0.73	62
Outgroup Attitudes												
Pretest	3.59	0.76	55	3.63	0.78	53	3.37	0.92	71	3.37	0.84	63
Posttest	3.47	0.82	54	3.51	0.82	53	3.25	0.84	67	3.30	0.80	61
Follow-up	3.18	0.70	46	3.57	0.74	53	3.14	0.85	61	3.19	0.81	62

Table 2. Intraclass correlation and class-level variance, multilevel modeling (Study 1)

	N	ICC	σ	SE	*p*
Contact quantity (Pretest)	243	.075	0.048	.017	.005
Contact quality (Pretest)	241	.038	0.027	.027	.321
Cross-group friendship (Pretest)	243	.021	0.013	.017	.449
Intergroup anxiety (Pretest)	243	.029	0.014	.015	.378
Outgroup attitudes (Pretest)	242	.044	0.030	.019	.113
Contact quantity (Posttest)	236	.055	0.035	.019	.071
Contact quality (Posttest)	235	.027	0.018	.022	.397
Cross-group friendship (Posttest)	236	.061	0.039	.023	.100
Intergroup anxiety (Posttest)	235	.027	0.013	.017	.442
Outgroup attitudes (Posttest)	235	.034	0.022	.020	.278
Contact quantity (Follow-up)	221	.096	0.058	.035	.098
Contact quality (Follow-up)	221	.009	0.004	.023	.868
Cross-group friendship (Follow-up)	222	.083	0.055	.027	.045
Intergroup anxiety (Follow-up)	222	.031	0.015	.013	.238
Outgroup attitudes (Follow-up)	222	.026	0.016	.014	.235

Note. ICC denotes intra-class correlation; SE denotes standard error.

Table 3. Intervention effects, multilevel regression model (Study 1)

	Decategorization-Categorization vs. Control			Categorization-Decategorization vs. Control			Categorization+ decategorization vs. Control		
	b	SE	*p*	*b*	SE	*p*	*b*	SE	*p*
Contact quantity (Pretest)	0.090	.146	.536	0.242	.207	.244	0.046	.177	.794
Contact quality (Pretest)	0.100	.145	.492	0.162	.125	.195	0.017	.196	.933
Cross-group friendship (Pretest)	0.213	.509	.675	0.258	.782	.741	-0.138	.493	.780
Intergroup anxiety (Pretest)	-0.027	.162	.866	-0.050	.172	.772	0.004	.187	.983
Outgroup attitudes (Pretest)	0.213	.138	.122	0.241	.167	.148	-0.018	.190	.924
Contact quantity (Posttest)	0.207	.112	.064	0.384	.146	.009	0.037	.147	.803
Contact quality (Posttest)	0.022	.103	.829	0.250	.099	.011	0.034	.186	.855
Cross-group friendship (Posttest)	0.146	.110	.184	0.361	.161	.025	0.087	.165	.597
Intergroup anxiety (Posttest)	-0.299	.134	.026	-0.331	.129	.010	-0.125	.155	.419

Outgroup attitudes (Posttest)	0.172	.093	.063	0.201	.152	.186	-0.054	.168	.748
Contact quantity (Follow-up)	0.194	.138	.161	0.473	.123	.000	-0.034	.176	.846
Contact quality (Follow-up)	0.062	.124	.619	0.288	.128	.025	-0.050	.133	.706
Cross-group friendship (Follow-up)	0.294	.165	.075	0.482	.153	.002	0.001	.148	.994
Intergroup anxiety (Follow-up)	-0.131	0.142	.355	-0.197	.119	.098	0.151	.129	.241
Outgroup attitudes (Follow-up)	-0.006	0.144	.966	0.379	.117	.001	-0.053	.136	.694

Table 4. Descriptives of study variables at posttest and follow-up in the experimental and control conditions (Study 2).

| | Posttest | | | | Follow-up | | | |
| | Categorization-Decategorization (*n=87*) | | Control group (*n=114*) | | Categorization-Decategorization (*n=74*) | | Control group (*n=90*) | |
Outcome variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Contact quantity	2.37	0.84	2.55	0.83	2.73	0.85	2.74	0.90
Contact quality	3.76	0.77	3.52	0.73	3.69	0.71	3.57	0.79
Cross-group friendships	2.72	1.00	2.56	0.92	2.80	1.01	2.75	1.01
Intergroup anxiety	2.15	0.70	2.22	0.68	2.23	0.68	2.35	0.75
Outgroup attitudes	3.62	0.68	3.31	0.69	3.41	0.74	3.26	0.80
Approach intentions	2.75	1.32	2.37	1.43	2.34	1.34	2.44	1.32
Avoid intentions	1.02	1.13	1.40	1.34	1.53	1.38	1.56	1.40

Table 5. Intraclass correlation and between-level variance, multilevel modeling (Study 2)

	N	ICC	σ	SE	p
Contact quantity (Posttest)	201	.045	.031	.018	.085
Contact quality (Posttest)	201	.029	.016	.012	.192
Cross-group friendship (Posttest)	201	.069	.062	.040	.118
Intergroup anxiety (Posttest)	201	.016	.007	.008	.366
Outgroup attitudes (Posttest)	201	.064	.031	.014	.030
Approach intentions (Posttest)	201	.021	.039	.049	.418
Avoid Intentions (Posttest)	201	.053	.082	.044	.062
Contact quantity (Follow-up)	164	.071	.054	.034	.113
Contact quality (Follow-up)	164	.031	.017	.024	.471
Cross-group friendship (Follow-up)	164	.215	.216	.070	.002
Intergroup anxiety (Follow-up)	164	.054	.027	.020	.172
Outgroup attitudes (Follow-up)	164	.047	.027	.026	.291
Approach intentions (Follow-up)	164	.049	.085	.082	.305
Avoid Intentions (Follow-up)	164	.010	.012	.037	.743

Table 6. Intervention effects, multilevel regression model (Study 2)

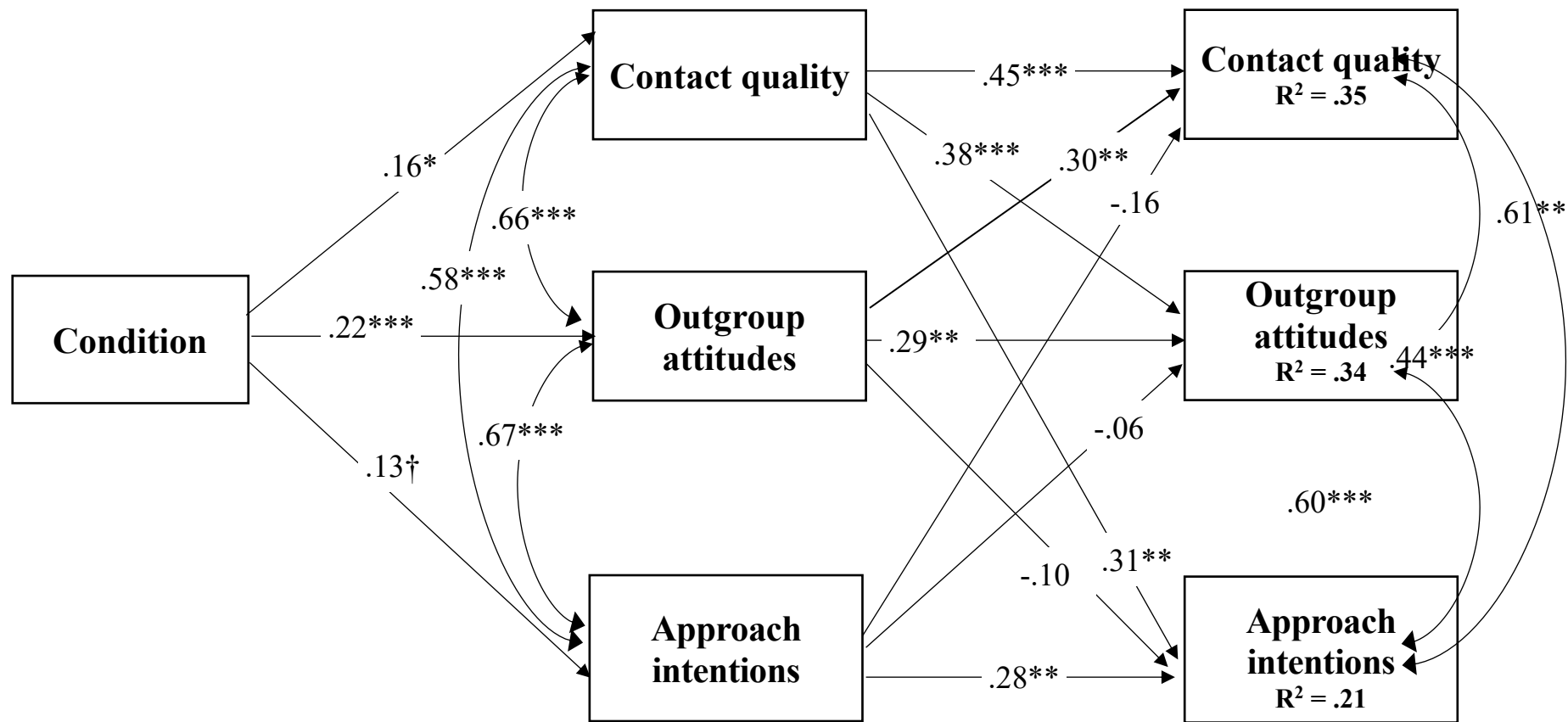|	Categorization/decategorization vs. control		
	b	SE	*p*
Contact quantity (Posttest)	0.163	.139	.239
Contact quality (Posttest)	0.244	.112	.029
Cross-group friendship (Posttest)	0.125	.170	.463
Intergroup anxiety (Posttest)	-0.069	.106	.512
Outgroup attitudes (Posttest)	0.312	.111	.005
Approach Intentions (Posttest)	0.372	.179	.038
Avoid Intentions (Posttest)	0.363	.196	.064
Contact quantity (Follow-up)	-0.013	.176	.940
Contact quality (Follow-up)	0.117	.127	.356
Cross-group friendship (Follow-up)	0.001	.272	.996
Intergroup anxiety (Follow-up)	-0.114	.130	.381
Outgroup attitudes (Follow-up)	0.139	.133	.297
Approach Intentions (Follow-up)	-0.134	.245	.584
Avoid Intentions (Follow-up)	-0.031	.224	.889

Table 7. Indirect effects of the experimental intervention on quality of contact, outgroup attitudes, and approach behavioral intentions at follow-up, bootstrap analysis (5,000 resamples; Study 2)

Mediator (Posttest)	Dependent Variable (Follow-up)	Estimate	BCa 95% bootstrap confidence interval	
			LLCI	ULCI
Contact quality	Contact quality	0.072	0.013	0.157
Contact quality	Outgroup attitudes	0.061	0.012	0.140
Contact quality	Approach intentions	0.050	0.008	0.125
Outgroup attitudes	Contact quality	0.066	0.018	0.145
Outgroup attitudes	Outgroup attitudes	0.066	0.014	0.149
Outgroup attitudes	Approach intentions	-0.099	-0.022	0.031
Approach intentions	Contact quality	-0.021	-0.076	0.003
Approach intentions	Outgroup attitudes	-0.008	-0.058	0.015
Approach intentions	Approach intentions	0.037	0.004	0.103

Note. BCa = bias-corrected and accelerated; LLCI = lower level of confidence interval; ULCI = upper level of confidence interval. Unstandardized parameters are reported.

Figure 1. Path analysis with observed variables testing delayed intervention-based effects; condition coded 1 for experimental condition, 0 for control condition (Study 2).

† $p < .06$, * $p < .05$, ** $p < .01$, *** $p \leq .001$

| INTERVENTION | Posttest | Follow-up |