Prejudice Reduction: What Works? A Review and Assessment of Research and Practice

Elizabeth Levy Paluck[1] and Donald P. Green[2]

[1]Harvard Academy for International and Area Studies, Weatherhead Center for International Affairs, Harvard University, Cambridge, Massachusetts 02138; email: epaluck@wcfia.harvard.edu

[2]Institution for Social and Policy Studies, Yale University, New Haven, Connecticut 06520-8209; email: donald.green@yale.edu

Key Words

field experiments, evaluation, stereotype reduction, cooperative learning, contact hypothesis, peace education, media and reading interventions, diversity training, cultural competence, multicultural education, antibias education, sensitivity training, cognitive training

Abstract

This article reviews the observational, laboratory, and field experimental literatures on interventions for reducing prejudice. Our review places special emphasis on assessing the methodological rigor of existing research, calling attention to problems of design and measurement that threaten both internal and external validity. Of the hundreds of studies we examine, a small fraction speak convincingly to the questions of whether, why, and under what conditions a given type of intervention works. We conclude that the causal effects of many widespread prejudice-reduction interventions, such as workplace diversity training and media campaigns, remain unknown. Although some inter-group contact and cooperation interventions appear promising, a much more rigorous and broad-ranging empirical assessment of prejudice-reduction strategies is needed to determine what works.

Contents

INTRODUCTION

By many standards, the psychological literature on prejudice ranks among the most impressive in all of social science. The sheer volume of scholarship is remarkable, reflecting decades of active scholarly investigation of the meaning, measurement, etiology, and consequences of prejudice. Few topics have attracted a greater

Prejudice: a negative bias toward a social category of people, with cognitive, affective, and behavioral components

range of theoretical perspectives. Theorizing has been accompanied by lively debates about the appropriate way to conceptualize and measure prejudice. The result is a rich array of measurement strategies and assessment tools.

The theoretical nuance and methodological sophistication of the prejudice literature are undeniable. Less clear is the stature of this literature when assessed in terms of the practical knowledge that it has generated. The study of prejudice attracts special attention because scholars seek to understand and remedy the social problems associated with prejudice, such as discrimination, inequality, and violence. Their aims are shared by policymakers, who spend billions of dollars annually on interventions aimed at prejudice reduction in schools, workplaces, neighborhoods, and regions beset by intergroup conflict. Given these practical objectives, it is natural to ask what has been learned about the most effective ways to reduce prejudice.

This review is not the first to pose this question. Previous reviews have summarized evidence within particular contexts (e.g., the laboratory: Wilder 1986; schools: Stephan 1999; cross-nationally: Pedersen et al. 2005), age groups (e.g., children: Aboud & Levy 2000), or for specific programs or theories (e.g., cooperative learning: Johnson & Johnson 1989; intergroup contact: Pettigrew & Tropp 2006; cultural competence training: Price et al. 2005). Other reviews cover abroad range of prejudice-reduction programs and the theories that underlie them (e.g., Oskamp 2000, Stephan & Stephan 2001).

Our review differs from prior reviews in three respects. First, the scope of our review is as broad as possible, encompassing both academic and nonacademic research. We augment the literature reviews of Oskamp (2000) and Stephan & Stephan (2001) with hundreds of additional studies. Second, our assessment of the prejudice literature has a decidedly methodological focus. Our aim is not simply to canvass existing hypotheses and findings but to assess the internal and external validity of the evidence. To what extent have studies established that

interventions reduce prejudice? To what extent do these findings generalize to other settings? Third, building on prior reviews that present methodological assessments of cultural competence (Kiselica & Maben 1999) and anti-homophobia (Stevenson 1988) program evaluations, our methodological assessment provides specific recommendations for enhancing the practical and theoretical value of prejudice reduction research.

Scope of the Review

We review interventions aimed at reducing prejudice, broadly defined. Our purview includes the reduction of negative attitudes toward one group (one academic definition of prejudice) and also the reduction of related phenomena like stereotyping, discrimination, intolerance, and negative emotions toward another group. For the sake of simplicity, we refer to all of these phenomena as "prejudice," but in our descriptions of individual interventions we use the same terms as the investigator.

By "prejudice reduction," we mean a causal pathway from some intervention to a reduced level of prejudice. Excluded, therefore, are studies that describe individual differences in prejudice, as these studies do not speak directly to the efficacy of specific interventions. Our concern with causality naturally leads us to place special emphasis on studies that use random assignment to evaluate programs, but our review also encompasses the large literature that uses nonexperimental methods.

Method

Over a five-year period ending in spring 2008, we searched for published and unpublished reports of interventions conducted with a stated intention of reducing prejudice or prejudice-related phenomena. We combed online databases of research literatures in psychology, sociology, education, medicine, policy studies, and organizational behavior, pairing primary search words "prejudice," "stereo-type," "discrimination," "bias," "racism," "homophobia," "hate," "tolerance," "reconciliation," "cultural competence/sensitivity," and "multicultural" with operative terms like "reduce," "program," "intervention," "modify," "education," "diversity training," "sensitize," and "cooperat*." To locate unpublished academic work, we posted requests on several organizations' email listservs, including the Society for Personality and Social Psychology and the American Evaluation Association, and we reviewed relevant conference proceedings. Lexis-Nexis and Google were used to locate nonacademic reports by nonprofit groups, government and nongovernmental agencies, and consulting firms that evaluate prejudice. We examined catalogues that advertise diversity programs to see if evaluations were mentioned or cited. Several evaluation consultants sent us material or spoke with us about their evaluation techniques.

Our search produced an immense database of 985 published and unpublished reports written by academics and nonacademics involved in research, practice, or both. The assembled body of work includes multicultural education, antibias instruction more generally, workplace diversity initiatives, dialogue groups, cooperative learning, moral and values education, intergroup contact, peace education, media interventions, reading interventions, intercultural and sensitivity training, cognitive training, and a host of miscellaneous techniques and interventions. The targets of these programs are racism, homophobia, ageism; antipathy toward ethnic, religious, national, and fictitious (experimental) groups; prejudice toward persons who are overweight, poor, or disabled; and attitudes toward diversity, reconciliation, and multiculturalism more generally. We excluded from our purview programs that addressed sex-based prejudice (the literature dealing with beliefs, attitudes, and behaviors toward women and men in general, as distinguished from gender-identity prejudices like homophobia). Sex-based inequality intersects with and reinforces other group-based prejudice (Jackman 1994, Pratto & Walker 2004),

Prejudice reduction: a causal pathway from an intervention (e.g., a peer conversation, a media program, an organizational policy, a law) to a reduced level of prejudice

but given the qualitatively different nature and the distinctive theoretical explanations for sex-based prejudice and inequality (Eagly & Mlednic 1994, Jackman 1994, Sidanius & Pratto 1999), we believe relevant interventions deserve their own review. The resulting database (available at **www.betsylevypaluck. com**) constitutes the most extensive list of published and unpublished prejudice-reduction reports assembled to date.

This sprawling body of research could be organized in many different ways. In order to focus attention on what kinds of valid conclusions may be drawn from this literature, we divide studies according to research design. This categorization scheme generates three groups: nonexperimental studies in the field, experimental studies in the laboratory, and experimental studies in the field. **Supplemental Table 1** (follow the **Supplemental Material link** from the Annual Reviews home page at **http://www.annualreviews.org**) provides a descriptive overview of the database according to this scheme. The database comprises 985 studies, of which 72% are published. Nearly two-thirds of all studies (60%) are nonexperimental, of which only 227 (38%) use a control group. The preponderance of nonexperimental studies is smaller when we look at published work; nevertheless, 55% of published studies of prejudice reduction use nonexperimental de-

FIELD VERSUS LABORATORY EXPERIMENTS

In an experimental design, units of observation (e.g., individuals, classrooms) are assigned at random to a treatment and to placebo or no-treatment conditions. Field experiments are randomized experiments that test the effects of real-world interventions in naturalistic settings, but the distinction between field and lab is often unclear. The laboratory can be the site of very realistic interventions, and conversely, artificial interventions may be tested in a nonlaboratory setting. When assessing the degree to which experiments qualify as field experiments, one must consider four aspects of the study: (*a*) participants, (*b*) the intervention and its target, (*c*) the obtrusiveness of intervention delivery, and (*d*) the assessed response to the intervention.

signs. Of the remaining studies, 284 (29%) are laboratory experiments and 107 (11%) are field experiments (see sidebar Field Versus Laboratory Experiments). A disproportionate percentage of field experiments are devoted to school-based interventions (88%).

Within each category, we group studies according to their theoretical approach or intervention technique, assessing findings in light of the research setting, participants, and outcome measurement. A narrative rather than a meta-analytic review suits this purpose, in the interest of presenting a richer description of the prejudice-reduction literature. Moreover, the methods, interventions, and dependent variables are so diverse that meta-analysis is potentially meaningless (Baumeister & Leary 1997; see also Hafer & Bègue 2005), especially given that many of the research designs used in this literature are prone to bias, rendering their findings unsuitable for meta-analysis.

Our review follows the classification structure of our database. We begin with an overview of nonexperimental prejudice-reduction field research. This literature illustrates not only the breadth of prejudice-reduction interventions, but also the methodological deficiencies that prevent studies from speaking authoritatively to the question of what causes reductions in prejudice. Next we turn to prejudice reduction in the scientific laboratory, where well-developed theories about prejudice reduction are tested with carefully controlled experiments. We examine the theories, intervention conditions, participants, and outcome measures to ask whether the findings support reliable causal inferences about prejudice reduction in nonlaboratory settings. We follow with a review of field experiments in order to assess the correspondence between these two bodies of research. Because field experiments have not previously been the focus of a research review, we describe these studies in detail and argue that field experimentation remains a promising but underutilized approach. We conclude with a summary of which theoretically driven interventions seem most promising in light of current evidence, and we provide recommendations for future

research (see sidebar Public Opinion Research and Prejudice Reduction).

NONEXPERIMENTAL RESEARCH IN THE FIELD

Random assignment ensures that participants who are "treated" with a prejudice-reduction intervention have the same expected background traits and levels of exposure to outside influences as participants in the control group. Outcomes in a randomized experiment are thus explained by a quantifiable combination of the intervention and random chance. By contrast, in nonexperimental research the outcomes can be explained by a combination of the intervention, random chance, and unmeasured pre-existing differences between comparison groups. So long as researchers remain uncertain about the nature and extent of these biases, nonexperimental research eventually ceases to be informative and experimental methodology becomes necessary to uncover the unbiased effect (Gerber et al. 2004). For these reasons, randomized experiments are the preferred method of evaluation when stakes are high (e.g., medical interventions).

Prejudice is cited as a cause of health, economic, and educational disparities (e.g., American Psychological Association 2001), as well as terrorism and mass murder (Sternberg 2003). For scientists who understand prejudice as a pandemic of the same magnitude as that of AIDS or cancer, a reliance on nonexperimental methods seems justifiable only as a short-run approach en route to experimental testing. Nevertheless, in schools, communities, organizations, government offices, media outlets, and health care settings, the overwhelming majority of prejudice-reduction interventions (77%, or 367 out of the 474 total field studies in our database) are evaluated solely with nonexperimental methods, when they are evaluated at all.

Studies with No Control Group

The majority of nonexperimental field studies do not use a control group to which an inter-

PUBLIC OPINION RESEARCH AND PREJUDICE REDUCTION

It is ironic but not coincidental that the largest empirical literature on the subject of prejudice—namely, public opinion research on the subject of race and politics—has little, if any, connection to the subject of prejudice reduction. Many of the most important and influential theories about prejudiced beliefs, attitudes, and actions have grown out of public opinion research. These theories examine the role of preadult socialization experiences (Sears 1988), group interests and identities (Bobo 1988), political culture and ideology (Sniderman & Piazza 1993), and mass media portrayals of issues and groups (Gilliam & Iyengar 2000, Mendelberg 2001). They diagnose the origins of prejudice, often tracing it to large-scale social forces such as intergroup competition for status and resources, but rarely do they propose or test interventions designed to ameliorate prejudice. Taking prejudice as a fixed personal attribute, this literature instead tends to offer suggestions about how to frame issues (e.g., public spending on welfare) in ways that mitigate the expression of prejudice (e.g., by reminding respondents that most welfare recipients are white).

vention group may be compared; most evaluations of sensitivity and cultural-competence programming, mass media campaigns, and diversity trainings are included in this category. Many no-control evaluations use a postintervention feedback questionnaire. For example, Dutch medical students described their experiences visiting patients of different ethnicities (van Wieringen et al. 2001), and Canadian citizens reported how much they noticed and liked the "We All Belong" television and newspaper campaign (Environics Research Group Limited 2001). Other feedback questionnaires ask participants to assess their own change: Diversity-training participants graded themselves on their knowledge about barriers to success for minorities and the effects of stereotypes and prejudice (Morris et al. 1996). Other no-control group studies use repeated measurement before and after the intervention: We were unable to locate a sensitivity- or diversity-training program for police that used more than a prepost survey of participating officers. Such strategies may reflect a lack of resources for,

understanding of, or commitment to rigorous evaluation.

Notwithstanding the frequency with which this repeated measures design is used, its defects are well known and potentially severe (Shadish et al. 2002). Change over time may be due to other events; self-reported change may reflect participants' greater familiarity with the questionnaire or the evaluation goals rather than a change in prejudice. Although such methodological points may be familiar to the point of cliché, these basic flaws cast doubt on studies of a majority of prejudice-reduction interventions, particularly those gauging prejudice reduction in medical, corporate, and law enforcement settings.

Qualitative Studies

A number of purely qualitative studies have recorded detailed observations of an intervention group over time with no nonintervention comparison (e.g., Roberts 2000). These studies are important for generating hypotheses and highlighting social psychological processes involved in program take-up, experience, and change processes, but they cannot reliably demonstrate the impact of a program. Qualitative measurement has no inherent connection to nonexperimental design, though the two are often conflated (e.g., Nagda & Zúñiga 2003, p. 112). Qualitative investigation can and should be used to develop research hypotheses and to augment experimental measurement of outcomes.

Cross-Sectional Studies

Diversity programs and community desegregation policies are often evaluated with a cross-sectional study. For example, one study reported that volunteer participants in a company's "Valuing Diversity" seminar were more culturally tolerant and positive about corporate diversity than were "control" employees—those who chose not to attend the seminar (Ellis & Sonnenfield 1994). Even defenders of diversity training would concede that peo-

ple with positive attitudes toward diversity are more likely to voluntarily attend a diversity seminar. Such evaluations conflate participants' predispositions with program impact. Although many cross-sectional studies report encouraging results, post hoc controls for participant predispositions cannot establish causality, even with advanced statistical techniques (Powers & Ellison 1995), due to the threat of unmeasured differences between treatment and control groups.

Quasi-Experimental Panel Studies

Prejudice-reduction interventions in educational settings, and some in counseling and diversity training, are more likely to receive attention from academically trained researchers who employ control groups and repeated measurement (e.g., Rudman et al. 2001). But with the exception of a few studies that use near-random assignment, most of these studies' findings have questionable internal validity.

For one, many quasi-experimental evaluations choose comparison groups that are substantially different from the intervention participants—such as younger students or students in a different school. Others choose comparison groups and assess preintervention differences more exactly. To evaluate a social justice educational program focused on dialogue and hands-on experience, investigators administered a pretest to all University of Michigan freshmen, some of whom had already signed up for the program (Gurin et al. 1999). Using this pretest, investigators selected a control group that was similar to program volunteers in gender, race/ethnicity, precollege and college residence, perspective taking, and complex thinking. After four years and four post-tests, results demonstrated that white students in the program were, among other things, more disposed to see commonality in interests and values with various groups of color than were white control students. This impressive study demonstrates the great lengths to which researchers must go to minimize concerns about selection bias, and yet no amount of

---

Qualitative studies: studies that gather narrative (textual, nonquantified) data and typically observe rather than manipulate variables

Cross-sectional study: design in which two or more naturally existing (i.e., not randomly assigned) groups are assessed and compared at a single time point

preintervention measurement can guarantee that the nonrandom treatment and control groups are equivalent when subjects self-select into the treatment group. Studies such as this one provide encouraging results that merit further testing using randomized designs (see also Rudman et al. 2001).

Near-Random Assignment

Fewer than a dozen studies have used comparison groups that were composed in an arbitrary, near-random fashion. Near-random assignment bolsters claims of causal impact insofar as exposure to the intervention is unlikely to be related to any characteristic of the intervention group. A good example is a waiting list design. In one of the few studies of corporate diversity training able to speak to causal impact (Hanover & Cellar 1998), a company's human resources department took advantage of a phased-in mandatory training policy and assigned white managers to diversity training or waiting list according to company scheduling demands. After participating in a series of sessions involving videos, role-plays, discussions, and anonymous feedback from employees in their charge, trainees were more likely than untrained managers to rate diversity practices as important and to report that they discourage prejudiced comments among employees. Unfortunately, all outcomes were self-reported, and managers may have exaggerated the influence of the training as a way to please company administration. Putting this important limitation aside, this research design represents a promising approach when policy dictates that all members of the target population must be treated.

Conclusion: Nonexperimental Research

That we find the nonexperimental literature to be less informative than others who have reviewed this literature (e.g., Stephan & Stephan 2001) does not mean this research is uninformative with respect to descriptive questions. These

studies yield a wealth of information about what kinds of programs are used with various populations, how they are implemented, which aspects engage participants, and the like. However, the nonexperimental literature cannot answer the question of "what works" to reduce prejudice in these real-world settings. Out of 207 quasi-experimental studies, fewer than twelve can be considered strongly suggestive of causal impact (or lack thereof). Unfortunately, the vast majority of real-world interventions—in schools, businesses, communities, hospitals, police stations, and media markets—have been studied with nonexperimental methods. We must therefore turn to experiments conducted in academic laboratories and in the field to learn about the causal impact of prejudice reduction interventions.

EXPERIMENTAL RESEARCH CONDUCTED IN THE LABORATORY

Academics studying prejudice reduction in the laboratory employ random assignment and base their interventions on theories of prejudice. Laboratory interventions using inter-group approaches aim at changing group interactions and group boundaries. Interventions using individual approaches target an individual's feelings, cognitions, and behaviors. Building on prior reviews (Crisp & Hewstone 2007, Hewstone 2000, Monteith et al. 1994, Wilder 1986), we describe an array of laboratory interventions and assess the extent to which these studies inform real-world prejudice-reduction efforts.

Intergroup Approaches

Prejudice-reduction strategies that take an intergroup approach are based on the general idea that peoples' perceptions and behaviors favor their own groups relative to others. Two major lines of thought have inspired techniques to address this in-group/out-group bias: the contact hypothesis (Allport 1954), which recommends exposure to members of the out-group under

Quasi-experimental studies: experiments with treatment and placebo or no-treatment conditions in which the units are not randomly assigned to conditions

Contact hypothesis: under positive conditions of equal status, shared goals, cooperation, and sanction by authority, interaction between two groups should lead to reduced prejudice

certain optimal conditions, and social identity and categorization theories (Miller & Brewer 1986, Tajfel 1970), which recommend interventions that break down or rearrange social boundaries.

Contact hypothesis. The contact hypothesis states that under optimal conditions of equal status, shared goals, authority sanction, and the absence of competition, interaction between two groups should lead to reduced prejudice (Pettigrew & Tropp 2006). Although there have been dozens of laboratory studies since Allport's original formulation of the hypothesis, among the most compelling are Cook's (1971, 1978) railroad studies. Cook simulated interracial workplace contact by hiring racially prejudiced white young adults to work on a railroad company management task with two "coworkers," a black and a white research confederate. Participants believed that they were working a real part-time job. Over the course of a month, the two confederates worked with participants under the optimal conditions of the contact hypothesis. At the end of the study, participants rated their black coworkers highly in attractiveness, likeability, and competence, a significant finding considering the study took place in 1960s in the American South. Several months later, participants also expressed less racial prejudice than controls expressed in an ostensibly unrelated questionnaire about race relations and race-relevant social policies. This exemplary piece of laboratory research employed a realistic intervention and tested its effects extensively and unobtrusively.

Social identity and categorization theories. Laboratory interventions guided by social identity and categorization theories address a variety of group prejudices, but often experimenters create new groups to study using the well-known minimal group paradigm (MGP; Tajfel 1970). Participants are sorted into two groups based on an irrelevant characteristic, such as the tendency to overestimate the number of dots on a screen (in actuality, assignment to the groups is random). Simple classification is of-

ten enough to create prejudice between these newly formed groups, but some researchers enhance in-group preference by having participants play group games or read positive information about their own group. In non-MGP studies, participants are reminded of a preexisting group identity, such as academic or political party affiliation. Once battle lines are drawn, these interventions use one of four kinds of strategies for reducing prejudice between the two groups: decategorization, recategorization, crossed categorization, and integration—each of which has generated a subsidiary theoretical literature (Crisp & Hewstone 2007).

In a decategorization approach, individual identity is emphasized over group identity through instruction or encouragement from the researcher. For example, participants in one study were less likely to favor their own (randomly assembled) group over the other group when the two groups worked cooperatively under instructions to focus on individuals (Bettencourt et al. 1992).

In recategorization research, participants are encouraged to think of people from different groups as part of one superordinate group using cues such as integrated seating, shirts of the same color (e.g., Gaertner & Dovidio 2000), or shared prizes (Gaertner et al. 1999). These studies have succeeded in encouraging members of minimal groups and political affiliation-based groups to favor their in-group less in terms of evaluation and rewards and to cooperate more with the out-group (Gaertner & Dovidio 2000).

Crossed categorization techniques (Crisp & Hewstone 1999) are based on the idea that prejudice is diminished when people in two opposing groups become aware that they share membership in a third group. Most commonly, prejudice against a novel group is diminished when it is crossed with another novel group category using the MGP (e.g., Brown & Turner 1979, Marcus-Newhall et al. 1993).

Integrative models (Gaertner & Dovidio 2000, Hornsey & Hogg 2000b) follow crossed categorization techniques with their strategy of preserving recognition of group differences

within a common group identity. In laboratory experiments, the common group identity is created by highlighting a superordinate identity (e.g., a university) without diminishing the value of identities constituting it (e.g., science and humanities students; Hornsey & Hogg 2000a) or by having two groups use their distinct areas of expertise to solve a task under equal status conditions (Dovidio et al. 1997).

All of these approaches achieve a measure of success in reducing prejudice as defined by preference for one's own group. Few laboratory interventions, especially those that use the MGP, target out-group derogation. The decategorization model has been criticized for its failure to extend this bias reduction toward the entire group (Rothbart & John 1985) and for submerging meaningful subgroup identities (Berry 1984). The integrative and crossed categorization models claim the most empirical and normative support, and have been used to bolster arguments for multicultural policies such as appreciating ethnic diversity under a common national identity (e.g., Brewer & Gaertner 2001, Hornsey & Hogg 2000b). Mixed findings from crossed categorization techniques may reflect varying definitions of in-group bias (Mullen et al. 2001), or the fact that these interventions change the perception of group boundaries but do not reduce out-group bias (Vescio et al. 2004).

Individual Approaches

Prejudice-reduction techniques aimed at individual phenomena such as feelings and cognitions are guided by a diverse set of theories that recommend a wide range of strategies, including instruction, expert opinion and norm information, manipulating accountability, consciousness-raising, and targeting personal identity, self-worth, or emotion.

Instruction. Ignorance has long been blamed as one of the roots of prejudice (Stephan & Stephan 1984), and the laboratory has been used to test different instructional solutions. Applied didactic techniques have been devel-

oped by researchers working with the U.S. military and with corporations sending employees overseas, teaching people how to interpret behaviors of different cultural and/or racial groups (e.g., Landis et al. 1976).

Other instruction techniques focus on ways to think, such as training in complex thinking and in statistical logic, with the hypothesis that this will help individuals avoid faulty group generalizations. These approaches claim modest success: After training, students are more likely to write positive stories about a picture depicting an interracial encounter, to report friendliness toward racial and ethnic out-groups (Gardiner 1972), and to avoid stereotyping fictitious characters presented in a vignette (Schaller et al. 1996).

Expert opinion and norm information. A body of social psychological research shows that prejudiced attitudes and behaviors are powerfully influenced by social norms (Crandall & Stangor 2005) and that under certain conditions people are persuaded by expert opinion (Kuklinski& Hurley 1996). Telling participants that experts believe personality is malleable (a position that undermines stereotyping; Levy et al. 1998) or that racial stereotyping is not normative for their peer group (Stangor et al. 2001; see also Monteith et al. 1996) reduces stereotyp-ing against stigmatized groups in the laboratory. More subtle manipulations designed to convey a tolerant social norm (e.g., an antiracism advertisement; GR Maio, SE Watt, M Hewstone, & KJ Rees, unpubl. manuscr.) seem to produce weaker effects.

Manipulating accountability. Theories emphasizing the irrationality of prejudice predict that asking people to provide concrete reasons for their prejudices should reduce them. Accountability interventions have succeeded in MGP studies, in which participants allocated more points to a fictitious out-group when they were required to justify their allocation amounts (Dobbs & Crano 2001). Students who believed they would be held accountable to peers for their evaluations of a Hispanic

---

Social norms: perceptions that are descriptive of what people are doing or prescriptive of what people should do (as a member of a group, an organization, or a society)

student involved in a school disciplinary case were also less likely to stereotype this student (Bodenhausen et al. 1994).

Consciousness-raising. Research on implicit prejudice proliferated following striking demonstrations that prejudiced attitudes and beliefs can operate without a person's awareness or endorsement (Devine 1989). A number of "(un)consciousness-raising" strategies (Banaji 2001, p. 136) aim to combat implicit prejudice through thought suppression, awareness, reconditioning, and control (see Blair 2002 for a review).

Instructions to suppress stereotypes (i.e., push them out of awareness) have had the opposite intended effect by increasing the accessibility of such stereotypes (Galinsky & Moskowitz 2000). For example, business students who watched diversity training videos instructing them to suppress negative thoughts about the elderly evaluated older job candidates less favorably than did students who did not receive suppression instructions (Kulik et al. 2000). Some evidence suggests that stereotype suppression does not lead inexorably to higher rates of stereotyping or prejudiced behavior (Monteith et al. 1998), particularly when suppression is coupled with mental retraining exercises (Kawakami et al. 2000a,b), but the overall pattern of findings suggests suppression is not an effective prejudice-reduction strategy.

Laboratory experiments have also tested the opposite strategy: encouraging awareness of memories, attitudes, or beliefs that relate to prejudice. For example, one intervention required students to remember a time when they treated an Asian person in a prejudiced manner (Son Hing et al. 2002). As predicted, students who previously scored high on an implicit prejudice test—by solving word fragments with the negative stereotypical Asian words "sly" and "short"—were more likely to feel guilt over this memory and to encourage funding for an Asian student association on a subsequent questionnaire.

Other laboratory interventions aim to recondition implicit attitudes and beliefs. Some

use classical conditioning techniques—pairing stigmatized groups with positive images and words—to improve college students' implicit stereotypes about the elderly, black Americans, and skinheads (Karpinski & Hilton 2001; Kawakami et al. 2000a,b; Olson & Fazio 2008). Presenting positive images of famous black people (e.g., Martin Luther King) and negative images of famous whites (e.g., Charles Manson) reduced implicit prejudice as measured by the Implicit Attitudes Test (IAT; Greenwald et al. 1998), but conscious attitudes remained unchanged (Dasgupta & Greenwald 2001, Wittenbrink et al. 2001). Other studies alter implicit attitudes and social distancing behaviors through approach-avoidance conditioning—i.e., by asking subjects to pull forward on a joystick when presented with words or faces representing a stigmatized group (Kawakami et al. 2007).

Targeting emotions. Psychologists contend that emotional states can influence the expressions of prejudice (e.g., E. Smith 1993), and some perspective-taking interventions encourage the perceiver to experience the target's emotions (Batson 1991). Writing an essay from the perspective of an elderly person decreased subsequent stereotypes about the elderly; writing an essay from the perspective of the opposite MGP group led to more positive ratings of the out-group's personality characteristics (Galinsky & Moskowitz 2000, Vescio et al. 2003). Instructions to be empathic when reading about everyday discrimination against blacks eliminated the difference between participants' evaluations of white and black Americans (Stephan & Finlay 1999). Similarly, instructions to "focus on your feelings" as opposed to thoughts when watching a video portraying anti-black discrimination increased desire to interact with blacks, an effect that was explained by a change in emotions toward blacks as a group (Esses & Dovidio 2002). This particular intervention did not change participants' beliefs or policy endorsements concerning blacks.

Targeting value consistency and self-worth.
Two related social psychological theories of motivation explain how the need to maintain consistency among valued cognitions or behaviors or to protect their self-worth might move people to express or repress prejudice. Festinger's cognitive dissonance theory (1957) has been used in several laboratory interventions that encourage participants to see prejudice as inconsistent with some valued attitude or trait. For example, college students were also more likely to soften pre-existing anti-black positions on social policies and to report more egalitarian attitudes and beliefs after agreeing to write public statements in favor of pro-black policies (Eisenstadt et al. 2003).

Steele's self-affirmation theory (1998) predicts that people will resist derogating others when their own self-worth is affirmed. Laboratory results are supportive: Individuals who affirmed their self-image by writing about their values or who received positive feedback about their intelligence were more likely to rate a Jewish job candidate positively in terms of her personality and her suitability for the job (Fein & Spencer 1997). Receiving positive feedback from a black manager of the laboratory experiment also decreased the amount of negative black stereotypes on a word-completion task (Sinclair & Kunda 1999).

Lessons for the Real World from Laboratory Experiments

Laboratory experiments test a wide range of prejudice reduction theories with a high degree of creativity and precision. Computers, video cameras, and even physiological measurements track manifestations of prejudice change. The laboratory environment and the experimental method lead to tight, internally valid conclusions about the causal impact of the intervention.

But do laboratory experiments yield reliable strategies for prejudice reduction in the world? Specifically, in the drive for simplification and abstraction, do laboratory experiments eliminate elements of their interventions, environ-ments, and theories that are critical to the external validity of their lessons for real-world prejudice reduction?

Interventions. Laboratory studies typically test quick fixes. Consider a typical minimal group paradigm experiment, in which prejudice is created, modified, and reassessed over the course of one hour. Brief manipulations can have powerful effects (e.g., Bargh et al. 1996), but studies rarely test to see if the change lasts longer than the study period.

Many laboratory prejudice interventions are also subtle; above we reviewed techniques based on slight changes in instructions, t-shirt color, and seating assignments. By contrast, real-world institutions are much more heavy-handed: They impose speech codes, citizenship requirements, immigration quotas, and economic sanctions that shape intergroup perceptions and relationships. Lessons on the power of authority and conformity handed down by Milgram, Asch, and Zimbardo have not been fully exploited in laboratory prejudice-reduction research. Two exceptions are research on conformity to perceived norms of prejudice (e.g., Stangor et al. 2001) and on orders to suppress stereotyping (e.g., Galinsky & Moskowitz 2000). Subtle manipulations undoubtedly have many advantages and applications, yet an exclusive focus on subtle techniques means that the laboratory is not approximating the full range of situational interventions.

A broader point is that laboratory interventions are often separated and abstracted from their real-world modalities. For example, in laboratory studies of empathy and prejudice reduction, participants receive instructions from the experimenter to imagine others' feelings. In the world, this message would be evoked within a moving speech, by a movie, or by the example of a peer. People interpret messages differently depending on who delivers the message and in what manner (Kuklinski & Hurley 1996). Laboratory studies eliminate larger institutions and social processes in which interventions are embedded—which may fundamentally change

Self-affirmation theory: predicts that when the self is under threat, people derogate others to affirm their self-identity; they refrain from other-derogation when their identity is affirmed

the impact and intervening psychological processes of the intervention.

Environment. Laboratory experiments themselves supply evidence challenging the external validity of the laboratory environment—to name a few, the presence of others affects emotional reactions (Ruiz-Belda et al. 2003), and a brief discussion with a peer can eliminate the influence of an authority's opinion (Druckman & Nelson 2003). The lack of correspondence between mundane living conditions and laboratory environments may be particularly damning for prejudice research, given some theoretical views that prejudice is a social norm set by peers and by the structure of the immediate situation (Crandall & Stangor 2005). Laboratory experiments like Cook's railroad job experiments address this concern by making the laboratory both an experimentally controlled and a realistic environment.

Populations. Warnings that North American college students differ from the general population (Sears 1986) are often acknowledged but disregarded by laboratory researchers. These students, who comprise the overwhelming majority of laboratory participants, are particularly exceptional when it comes to expressions of prejudice. At least in the United States, college students report less prejudice than does the average individual (Judd et al. 1995) and are more aware of social proscriptions against the expression of prejudice (Crandall et al. 2002). College subjects come to the lab having had more exposure to some form of diversity or antibias training (McCauley et al. 2000).

Prejudices. If prejudice were likened to a sickness, many laboratory interventions would be walk-in clinics, built to handle low-grade prejudices. Many studies get around the problem of college students' politically correct response patterns by studying socially acceptable prejudices against skinheads, political parties, or the elderly (e.g., Karpinski & Hilton 2001). Moreover, prejudices created with the minimal group paradigm for maximum experimental control

lack the historical, political, and economic forces that animate and sustain real-world prejudice, and ". . .a fundamental challenge remains to discover ways of changing 'hard-core' prejudiced beliefs" (Monteith et al. 1994).

Outcome measures. Measuring prejudice is a formidable challenge for all types of research, including laboratory studies. Behaviors measured in the laboratory are often low-stakes abstractions of real-world behaviors, such as giving up tokens to another group or brief interactions with a stranger. Laboratory investigators also rely on indirect measures to measure racial and ethnic prejudice. The linguistic bias index is an indirect measure in which verbs and nouns from participants' writing samples are classified according to their implication that out-group failings are dispositional while in-group failings are situation-specific (Maass 1999). Other measures gauge subtle forms of unease or reticence more than antipathy. One example is "immediacy behaviors," such as physical posture toward and distance from another person (Kawakami et al. 2007).

Controversy surrounds the interpretation of a "prejudiced score" on tests of implicit prejudice such as the IAT. Some studies find implicit prejudice to be correlated with the disintegration of real-world interracial friendships (Towles-Schwen & Fazio 2006), but a recent meta-analysis found that across 32 studies the IAT's ability to predict discriminatory behavior varies widely and sometimes inexplicably (AG Greenwald, TA Poehlman, E Uhlmann, & MR Banaji, unpubl. manuscr.). Other measures of implicit prejudice, such as word fragment completion (e.g., "short" versus "smart" in the case of Asians; Son Hing et al. 2002), are not empirically linked to behavior. Most importantly, few studies have connected the reduction of implicit prejudice with a reduction in prejudiced behavior.

Theories. A thorough review of theories developed in the laboratory goes beyond the scope of this essay, but we note that theory development in the laboratory mostly takes its

lead from other laboratory experiments. We worry this creates a theoretical echo chamber in which ideas are not cross-fertilized by research conducted in real-world settings. Additionally, most theory developed in laboratories addresses one or two dimensions of prejudice, (e.g., cognition and behavior); one may question whether these theories are sufficiently multifaceted to predict how and when prejudice is expressed or changed in real-world settings (Paluck 2008).

The ultimate arbiters of the debate about the external validity of prejudice-reduction laboratory studies are research programs that straddle the two settings. Currently, such programs are extremely rare. An exception is the cooperative learning research program (e.g., Johnson & Johnson 1989, Roseth et al. 2008), in which field studies are sometimes inconsistent with laboratory results (e.g., Rich et al. 1995). One research program hardly settles the issue, and the correspondence between findings in the lab and field merits active investigation.

Conclusion: Experimental Research in the Laboratory

Reviewers of the psychological prejudice-reduction literature regularly comment that "... promising laboratory studies always need to be tested in field settings" (Miller & Harrington 1990, p. 218), but translation is rarely attempted, and psychologists frequently offer their laboratory findings as guidance for policymakers (e.g., Crisp & Hewstone 2007, p. 239). Those interested in creating effective prejudice- reduction programs must remain skeptical of the recommendations of laboratory experiments until they are supported by research of the same degree of rigor outside of the laboratory.

EXPERIMENTAL RESEARCH CONDUCTED IN THE FIELD

Over a half-century ago, psychologist Stuart Cook endeavored to make his research"... both socially useful and scientifically meaningful"

(Selltiz & Cook 1948) by using lab and survey methods to develop the theoretical models he then tested using "true experiments" in the field (Cook 1985, p. 452). To what extent have prejudice-reduction researchers followed this example?

Of the hundreds of reports culled from our literature search, we identified 107 randomized field experiments. Thirty-six of these were studies of cooperative learning, which means that 71 experiments speak to the efficacy of all other types of prejudice interventions. To put this number into perspective, a PsychInfo database search for studies of one type of prejudice—implicit—retrieves 116 empirical articles. Our review's database contains four times as many laboratory experiments and five times as many nonexperimental field studies as noncooperative learning field experiments; this group of 71 studies is further dwarfed by the hundreds and perhaps thousands of unevaluated an-tiprejudice interventions implemented yearly in schools, businesses, and governments. Because the cooperative learning experiments have been summarized elsewhere (Roseth et al. 2008), **Supplemental Table 1** is confined to the 71 remaining studies.

Supplemental Table 1 describes these 71 field experiments, from the earliest in 1958 to present. Eighty percent of the studies are from North America. Almost one-third of these studies address prejudice against African Americans, 20% address multiple prejudices or are more generally "antibias" treatments; 13% of the studies address a non-African American group prejudice, including Mexican and Native Americans; 11% of the studies address "cultural competence"—comfort and ability to interact with people of different cultures. Of the remaining 18% of studies, 6 address prejudice against the disabled, 3 address prejudices against immigrants or refugees, 3 address religious prejudice, and 1 addresses prejudice against gay men.

Fifty-six percent of the interventions lasted one day or less. Excluding the cooperative learning studies, 84% of intervention studies took place with students or school

personnel. This means antiprejudice education has developed a research literature, whereas the rest of the prejudice-reduction enterprise lacks randomized controlled evaluations.

Evaluations also focus on volunteers (e.g., Haring 1987, Pagtolun-an & Clair 1986, Stewart et al. 2003). It is easy to understand why, for practical reasons, interventions would tend to be directed toward people who are open to their messages. Unfortunately, field research on prejudice reduction does not have much to say about influencing those who do not sign up for antiprejudice interventions. Four studies took place in settings of extreme intergroup conflict, measuring reactions to peace education, a media program, and diversity training in Israel, Rwanda, and South Africa, respectively. The literature provides little empirical guidance to policymakers seeking to intervene with populations living in conflict or postconflict environments.

The breadth of answers to the question "What reduces prejudice in the world?" narrows further when we probe these studies' designs. Several suffer from weak outcome measurement. Most rely solely on self-report questionnaires; only 11 studies involve directly observed measures of behavior (two gather third-party reports). We would expect behavioral measurement to be the strength of field studies, which take place in environments where the behaviors of interest actually occur. Many clever unobtrusive measures of real-world behavior have been developed (Crosby et al. 1980), but these measurement techniques are rarely used in this literature. One of the few exceptions is a study of a disability awareness program that used audit study methods, sending disabled and nondisabled confederates to ask for help from employees who had attended the program (Wikfors 1998).

Inadequate power is another frequent problem; approximately half of the studies have sample sizes of below 100 individuals. Thirteen of the studies with larger sample sizes assign groups (e.g., classrooms, schools) to treatment and control groups but fail to make the necessary corrections for intracluster correla-

tions within groups when calculating significance levels.

We now review the best of prejudice-reduction interventions and theories tested with field experiments. The most frequently studied interventions are cooperative learning (34% of all field experiments), entertainment (reading and media: 28%), discussion and peer influence (16%), and instruction (15%). We also review interventions that receive a great deal of attention in the lab but seldom in field settings: contact (10% of field experimental studies), cognitive training (5%), value consistency and self-worth interventions (4%), and social categorization (2%).

Cooperative Learning

Derived from Deutsch's (1949) theory of social interdependence and best known through Eliot Aronson's "Jigsaw classroom" technique (Aronson et al. 1978), cooperative learning lessons are engineered so that students must teach and learn from one another. For example, teachers in Jigsaw classrooms give each student one piece of the lesson plan, so that good lesson comprehension requires students to put together the pieces of the "puzzle" collectively. Approximately eight variants on this basic cooperative learning model exist (Slavin et al. 1984). Expected outcomes include interpersonal attraction, perspective taking, social support, and constructive management of conflict. Meta-analyses of the effects of cooperative techniques (which included nonexperimental results) on relationships crossing ethnic, racial, and ability boundaries have consistently confirmed a positive impact of cooperation on outcomes such as positive peer relationships and helpfulness (Johnson & Johnson 1989, Roseth et al. 2008). The few studies that investigate generalization of cross-group friendships to individuals outside of the immediate classroom find weaker effects (cf., Warring et al. 1985). Fewer studies measure generalization to the entire racial or ethnic group or track long-term effects. Nevertheless, the cooperative learning literature sets the standard

for programmatic field research on causal mechanisms. That 79% of all U.S. elementary schools by the early 1990s used cooperative learning (Puma et al. 1993) attests to the influence a well-documented causal effect can have on policy implementation.

Entertainment

Books, radio, television, and film are vivid and popular couriers of many kinds of social and political messages. *Uncle Tom's Cabin*, published by Harriet Beecher Stowe in 1852, was heralded as the turning point in American abolitionist opinion—not only for the information it provided about the brutality of slavery, but also for its ability to "go to the heart" (cited in Strange 2002, p. 263).

Reading and media interventions, most of them using an engaging narrative rather than an informational style, comprise 42% of all non-cooperative learning prejudice-reduction field experiments. We analyze the reading interventions separately because they share the specific modality of a book, but all of these interventions potentially draw from many of the same change processes via narrative persuasion or extended contact, which we describe below.

Reading. All 17 field experiments on the impact of reading on prejudice were conducted in schools—studies have yet to examine the effect of literature on prejudice among general audiences. One clear advantage of these reading experiments is that they evaluate substantially longer interventions compared to other field interventions. Whereas half of all field experiments focused on an intervention lasting one day or less, reading interventions lasted five weeks on average. Children in pre-K through high school were randomly assigned to read stories from or about other cultures (Gwinn 1998; Wham et al. 1996) about African, Native American, or disabled people (Clunies-Ross & O'Meara 1989, Fisher 1968, Hughes 2007, Yawkey 1973), or about contact between children from different groups (Cameron & Rutland 2006, Cameron et al. 2006, Liebkind & McAlister 1999, Slone et al. 2000).

Eleven of the 17 field experiments on reading report positive results, mostly for self-reported attitudinal outcomes; none measured behavior. The evidence is mixed or null for multicultural literature, more positive for portrayals of people of another culture or race, and wholly positive for books that portray contact between children who are similar to the audience and children of different cultures or races. For example, Cameron & Rutland (2006) randomly assigned 253 five- to eleven-year-old English schoolchildren to listen to stories about a nondisabled child's close friendship with a disabled child. The books described the two children's adventures, such as exploring in the woods. Across the three randomized conditions, the books emphasized characters' individual characteristics versus their group membership, versus a different unrelated story. Like eight other reading field experiments, this intervention included a group discussion led by the experimenter at the end of the story. Story hours took place once per week for six weeks.

The positive attitudinal effects found in this study and in four others that examined stories about intergroup friendship are consistent with the positive impact of vicarious experiences of cross-group friendship that is predicted by the extended contact hypothesis (Wright et al. 1997). Theories of narrative persuasion suggest additional processes that could explain prejudice reduction findings from reading field experiments that were not as theoretically motivated. For example, stories are channels for communicating social norms—descriptions of what peers are doing (and therefore what the reader or listener should do; Bandura 1986, 2006). Narratives encourage perspective taking (Strange 2002) and empathy (Zillmann 1991); texts can "transport" us into an imaginative world where we inhabit other characters, learn new things, and in general remove filters that might otherwise screen out different perspectives (Gerrig 1993, Green & Brock 2002).

Media. We found 13 media studies, 7 of which were one-time viewing experiences, such as a

documentary or an educational movie. Because few programs were based on theory, it is difficult to draw broad lessons from the pattern of their findings, but like many of the reading studies, their results are suggestive for those interested in narrative persuasion, empathy, perspective taking, social norms, and the like. Most media experiments were conducted in schools, on media-driven multicultural or antibias education. Few have gauged the impact of media on large audiences or the impact of large-scale media campaigns (which span long periods of time or multiple theatres, cable networks, or airwaves). Two exceptions are a study of a children's multicultural television series (Mays et al. 1973) and of a "reconciliation" radio soap opera.

A year-long field experiment in Rwanda (Paluck 2008, Paluck & Green 2008) tested the impact of a radio soap opera featuring a fictional story of two Rwandan communities and their struggles with prejudice and violence. The program aimed to change beliefs using didactic messages and to influence perceived norms through realistic radio characters who could speak to audience experiences. Nearly 600 Rwandan citizens, prisoners, and genocide survivors listened to the program or to a health radio soap opera. The investigators found the radio program affected listeners' perceptions of social norms and their behaviors with respect to intermarriage, open dissent, cooperation, and trauma healing, but did little to change listeners' personal beliefs. The program also encouraged greater empathy. The results pointed to an integrated model of behavioral prejudice reduction in which intergroup behaviors are linked more closely to social norms than to personal beliefs.

Discussion is featured in many studies of entertainment, because storytelling and media consumption are inherently social practices, and also because an implicit theory in much intervention design is that peer discussion amplifies message impact. We now turn to consider discussion and peer influence as interventions in their own right and how they have fared in field experimental research.

Discussion and Peer Influence

Although psychologists examine group discussion for processes related to polarization of attitudes and minority influence, they seldom focus on communication for prejudice reduction. One of the few exceptions is Fisher (1968), which found that the addition of discussion strengthened the positive attitudinal effects of a reading intervention.

Evidence of the benefits (and potential pitfalls) of discussing opinions about inter-group relations is also found in peer influence studies. For example, Blanchard and colleagues (1991, 1994) find that white university females' opinions about a racial incident on campus conformed to the publicly expressed opinions of confederates who were randomly assigned to condone, condemn, or remain neutral in their reactions. Another study of norms, a field experiment assessing the Anti-Defamation League Peer Training program (Paluck 2006b), showed that students were able to influence close friends and casual acquaintances in their school with public behaviors such as speaking out against biased jokes. Although few field experiments have experimentally isolated the effects of normative communication and discussion in field interventions of prejudice reduction, these findings indicate that theories of social norms and mechanisms of small group or peer discussion are promising avenues for research and intervention.

Instruction

Under the umbrella category of "instruction" we find myriad interventions: multicultural education, "ethnic studies," stand-alone lectures, awareness workshops, and peace education. Few instructional techniques are guided by theoretical models of learning or prejudice reduction (see Bigler 1999 critiquing multicultural education in particular). The lack of theory may explain in part the lack of impressive findings. One notable exception is Lustig's (2003) investigation of a peace education program in Israel that aimed to increase perspective taking and

empathy using instruction about foreign conflicts. Twelfth-grade Israeli Jewish students were randomly assigned to a "permanent peace" curriculum (versus no curriculum) about conflicts in ancient Greece and modern-day Ireland. Questionnaire-based opinions about the Israeli-Palestinian conflict revealed no effect of the curriculum, but there were striking differences between essays students were asked to write from the Palestinian point of view. Curriculum student essays were more likely to be written in the first as opposed to the third person, and they were more sympathetic to damages and to the symmetry of Israeli-Palestinian conflict. This study is an excellent example of the benefits of multiple and non-traditional outcome measurement, and of interventions informed by theories of prejudice reduction.

Less-Frequently Studied Approaches in the Field

Given the academic focus on the contact hypothesis, social identity theory and related social categorization strategies, cognitive forms of prejudice, and motivational theories of identity and dissonance, the field experimental literature on these areas is surprisingly thin.

Contact hypothesis. What is most notable about field experiments categorized as "contact" experiments is their general lack of resemblance to the conditions of contact specified by Allport (1954). Within the small body of field experiments on contact, there is also a tendency to address prejudices that may be more related to unfamiliarity (e.g., disability) than to antipathy.

Among more recent studies, we find two of note. One study capitalized on random assignment of minority and white students to college dorm rooms (Duncan et al. 2003). The experiment's findings from a subsequent Internet-based survey are important for their suggestion that cross-race contact affects more general attitudes such as support for affirmative action, although weak effects on other attitudinal and be-

havioral outcomes suggest this finding requires more study. The second study, conducted with the Outward Bound camping expedition organization, randomly assigned 54 white teenagers to racially homogeneous (all white) or heterogeneous expedition groups (Green & Wong 2008). In these expeditions, an experienced leader teaches campers group survival techniques under most of Allport's (1954) conditions for ideal intergroup contact: equal status, a common (survival) goal, authority sanction, and intimate contact. One month after the two-to three-week trip, in an ostensibly unrelated phone survey, white teens from the heterogeneous groups reported significantly less aversion to blacks and gays and described themselves as less "prejudiced" compared to the homogeneous group teens. The intensity and naturalistic quality of the intervention recall the seminal field study of contact, the Sherif et al. (1961) Robbers Cave experiment. The study's limitations—small sample size, the lack of behavioral outcome measures, and short-term follow-up—invite replication and extension.

Social identity and categorization. Although principles of social identity and categorization theory broadly inform some field interventions (e.g., Cameron & Rutland 2006), very few field experiments have been designed to test crossed, integrated, re- and decategorization strategies developed in the laboratory. Two exceptions are studies by Nier et al. (2001) and Houlette et al. (2004) testing the Common Ingroup Identity model.

Over the course of 12 hour-long sessions, instructors in 61 randomly assigned first- and second-grade classrooms led discussions about sex, race, and body size exclusion from the "green circle of community" (Houlette et al. 2004), versus an enhanced program, versus no program. In the enhanced program, the perimeter of the classroom was circled with green tape and all students wore the same green vests. Mixed, modest results showed that children in the program classrooms were slightly more likely to favor drawings of cross-sex or -race children. Weight remained a powerful

predictor of children's hypothetical choices of playmates. The enhanced program did not amplify these effects (which speaks to our previous question about the real-world effects of subtle laboratory interventions such as seating arrangements and similar clothing).

Value consistency and self-worth. Compared to their importance in the laboratory literature, studies of the motivating forces of consistency and self-worth are scarce in the field experimental literature. A notable exception is the Rokeach value confrontation technique (Gray & Ashmore 1975, Rokeach 1971). Rokeach (1971, 1973) lectured college students about (fictitious) research findings on values revealing people who value equality are more likely to be sympathetic toward black Americans' civil rights (during this historical period, most North American students favored equality but not black civil rights). In postinterven-tion questionnaires that stretched as far as 17 months later, students from the lecture and the no-lecture classes increased their support of black civil rights, perhaps in part through exposure to the more liberal college atmosphere, but treatment students eventually outpaced others in their support (Rokeach never corrected for intraclass correlations, which should lead to more caution about the statistical significance of his findings). Twenty-one months later, twice as many experimental as control subjects were enrolled in an ethnic core course, and three to five months after the intervention, 51 treatment versus 18 control participants responded to solicitations sent by the NAACP (although a com-parable number of control students responded the following year). Although the strength of these results is at times mixed, this series of studies is notable for its behavioral measures and longitudinal design.

Cognitive training. Excellent laboratory and quasi-field experimental research has examined stereotype retraining with young children (Levy 1999, Levy et al. 2004), but there are very few studies of cognitive retraining in the field experimental literature. Five field experi-

ments, all conducted on North American students, show weak results in both the short- and long-term (e.g., Katz 1978, 2000).

Lessons of Field Experimental Research

The strongest conclusion to be drawn from the field experimental literature on prejudice reduction concerns the dearth of evidence for most prejudice-reduction programs. Few programs originating in scientific laboratories, nonprofit or educational organizations, government bureaus, and consulting firms have been evaluated rigorously. Theories with the strongest support from the laboratory sometimes receive scant attention in the field. Entire genres of prejudice-reduction interventions, including moral education, organizational diversity training, advertising, and cultural competence in the health and law enforcement professions, have never been tested, as well as countless individual programs within the broad genre of educational interventions.

Nonetheless, the field experimental literature on prejudice reduction suggests some tentative conclusions and promising avenues for reducing prejudice. Cooperative learning emerges as an important tool for breaking down boundaries between students. This research program should be emulated and extended. More research is needed on the behavioral and longitudinal impact of cooperative learning and its impact on out-group dislike as well as in-group preference.

Media and reading interventions bear out assorted predictions of the extended contact hypothesis and of narrative persuasion, specifically that extended contact can reduce out-group hostility, and narratives can communicate norms and inspire empathy and perspective taking. Theoretically driven programs of research on entertainment and narrative interventions would systematize what is at present a rather disjointed set of studies and findings. Extended contact and narrative persuasion might also provide frameworks for other strategies associated with empathy or perspective taking

such as role playing (e.g., the Jane Elliot Blue Eyes/Brown Eyes intervention; Stewart et al. 2003), which have met with mixed success, perhaps in part because of a lack of theoretical grounding.

Given the importance of social psychological processes such as obedience and conformity, experimental evaluations of peer influence and discussion should become a priority for future field research. Isolating the influence of discussion from the impact of the intervention itself is an important future step (Kelman & Fisher 2003, p. 335).

Recommendations

Few rigorous field studies to date have addressed psychology's most important theories of interpersonal and intrapersonal prejudice change: contact, social identity and categorization, identity and value-motivated techniques, and social cognitive (stereotype and implicit prejudice) interventions. We recommend more field experimentation on social psychology's principal theories of prejudice.

The strength of field experimentation rests not only in its ability to assess causal relationships but also in its ability to assess whether an intervention's effects emerge and endure among the cacophony of real-world influences including larger political and economic changes and proximal social pressures and distractions. We recommend that more field experiments assess the strength and persistence of effects with outcome measurement that moves beyond the site of the intervention. Types of outcome measures should be increased to capture prejudice from different angles, especially with unobtrusive and behavioral measures, and the settings should be expanded so as to augment our knowledge about changing prejudice outside of the classroom and with older populations.

Although laboratory studies concentrate on interventions targeted at specific forms of prejudice (e.g., stereotyping), the complexities of real-world contexts often force the field experimentalist to design and parse the impact of multidimensional interventions aimed at sev-

eral forms of prejudiced speech, behavior, and attitudes. Studying prejudice reduction in the field opens our eyes to the utility of more multidimensional theories of prejudice reduction. Field experimentation can be productive for assessing the functional interdependence of cognitive, affective, normative, and other forms of prejudice, and thus for building prejudice-reduction theories based on this recognition of the interrelationships and on the sequencing and long-term effects of change in one part of the system (e.g., an intervention that changes social norms, which then affect behaviors and finally beliefs; see Paluck 2008 for one such attempt). Field experimentation is not only a method for testing theoretical ideas developed in the laboratory—the field itself should be used as a laboratory for generating richer, more multidimensioned theory.

DISCUSSION

In terms of size, breadth, and vitality, the prejudice literature has few rivals. Thousands of researchers from an array of disciplines have addressed the meaning, measurement, and expression of prejudice. The result is a literature teeming with ideas about the causes of prejudice. In quantitative terms, the literature on prejudice reduction is vast, but a survey of this literature reveals a paucity of research that supports internally valid inferences and externally valid generalization.

In order to formulate policies about how to reduce prejudice, one currently must extrapolate well beyond the data, using theoretical presuppositions to fill in the empirical blanks. One can argue that diversity training workshops succeed because they break down stereotypes and encourage empathy. Alternatively, one can argue that such workshops reinforce stereotypes and elicit reactance among the most prejudiced participants. Neither of these conflicting arguments is backed by the type of evidence that would convince a skeptic. We currently do not know whether a wide range of programs and policies tend to work on average, and we are quite far from having an empirically grounded

Table 1   Summary of prejudice-reduction approaches, theories, and future directions for research

Intervention approach	Theoretical frameworks	Evidence needed
Supported by experimental evidence from field and laboratory		
Cooperative learning	Social Interdependence Theory	Longitudinal, generalization to wider groups, reduction of negative out-group attitudes
Entertainment	Extended contact, narrative persuasion (empathy, perspective taking, transport-imagery), social norm theory, social cognitive theory	Theory-driven programmatic research; studies of longer duration and with adults
Peer influence, discussion/dialogue	Social norm theory, small group influence, social impact theory, contact hypothesis	Field experimental evidence; isolation of effects of discussion from other aspects of intervention
Contact	Contact and extended contact hypothesis	Field experimental evidence for differing contact conditions and more antagonistic groups
Value consistency and self-worth	Cognitive dissonance, self-affirmation and self-perception theory	Field experimental evidence; evidence with "unmotivated" populations
Cross-cultural/intercultural training	Acculturation theory, Bhawuk/Landis model	Field experimental evidence; behavioral, longitudinal effects
Supported mostly by		
Social categorization	Social identity theory, crossed-categorization, common in-group identity, de- and recategorization	Field experimental evidence; evidence with antagonistic groups and longitudinal effects
Cognitive training	Implicit prejudice, classical conditioning	Field experimental evidence; longitudinal effects
In need of theoretical and		
Diversity training	Dependent on technique/modality used	Theory-driven intervention design and field experimentation
Multicultural, antibias, moral education	Socialization theories of prejudice, cognitive, moral development and learning theories	Field experimental evaluations with longitudinal outcome measurement
Sensitivity, cultural competence for health and law	Dependent on technique/modality used	Theory-driven intervention design and field experimentation
Conflict resolution	Interactive conflict resolution models	Theory-driven field experimentation

understanding of the conditions under which these programs work best.

Looking across all of the settings, populations, and methodologies used to study the reduction of prejudice, we classify the main approaches to prejudice reduction according to the evidence accumulated thus far for their impact in the real world, and we list theories and methods that could point the way forward (see **Table 1**).

Cooperative learning is the most outstanding example of theoretically driven, programmatic laboratory and field research; we hope future research will address questions about the longevity and generality of cooperative learning's effects. Although media, reading, and other forms of narrative and normative communication are not currently considered cutting-edge approaches, we point to the apparent success of this technique in the real world and to its potential for reducing prejudice through narrative persuasion, social norms, empathy, perspective taking, and extended contact. The persuasive and positive influence of peers (indirectly via observation or directly via discussion) is a promising area of prejudice reduction supported by laboratory research (Stangor et al. 2001) and by creative real-world interventions (Aboud & Doyle 1996; Blanchard et al. 1991, 1994; McAlister et al. 2000; Nagda et al. 2004; Paluck 2006b) highlighting the communicative and normative nature of prejudice change.

The contact hypothesis, which benefited from early and innovative field and laboratory studies, remains unproven in the real world due to the limited number of randomized studies conducted in field settings and the narrow range of prejudices tested in those studies. Researchers should aspire to extend real-world experimental tests to domains such as summer camps, multinational peacekeeping units, and refugee settlements. Other approaches that require more field experimental tests are consistency and self-worth interventions based on balance and self-affirmation theories, as well as cross-cultural training approaches. Given that motivation is a critical lever of change for these interventions, field tests would illuminate whether these techniques are successful with participants who are unmotivated to change, and what adjustments are needed in order to reach this population. Interventions aimed at changing cognitions (e.g., stereotypes or automatic associations) or cognitive abilities (e.g., complex thinking or statistical reasoning) have successfully reduced prejudice in the laboratory, but the magnitude and persistence of these effects also await testing in real-world settings.

Several areas of prejudice reduction are in need of research and theory. Although antibias, multicultural, and moral education are popular approaches, they have not been examined with a great deal of rigor, and many applications are theoretically ungrounded. Spending on corporate diversity training in the United States alone costs an estimated $8 billion annually (cited in Hansen 2003), and yet the impact of diversity training remains largely unknown (Paluck 2006a). Despite research showing that medical practitioners' negative bias can affect their administration of care (Flores et al. 2000) and reports of sharply increased demand within the law enforcement field following September 2001 (*New York Times*, Jan. 23, 2005), sensitivity trainings administered to medical personnel and police are rarely based on theory or subjected to rigorous evaluation. Finally, although there is a distinguished tradition of psychological research on conflict resolution for elite negotiators (Kelman & Fisher 2003), there is lit-

tle sustained experimental evaluation of conflict negotiation and reduction for the many millions of ordinary citizens living in conflict or postconflict settings (G Salomon & B Nevo, unpubl. manuscr.; cf., Bargal & Bar 1992).

Final Thoughts

Field experiments present a range of practical challenges, but we believe that the failure to implement field experimental designs is in part a failure of creativity. Random assignment to waiting lists solves the problem of control groups who wish to undergo treatment and represents a low-cost opportunity for randomized field experimentation. Randomly phasing in a program to different parts of a target population solves the problem of the "saturation model" intervention. For interventions where it is absolutely impossible to leave out a control group, researchers can use rigorous and underappreciated quasi-experimental techniques such as regression discontinuity (Shadish et al. 2002). A lack of field experimental training among practitioners who evaluate prejudicereduction programs, doubts about the feasibility of randomized field methodology, and insufficient incentives for academics to conduct "applied" research all contribute to the scarcity of randomized field experiments in prejudice reduction. We believe that each of these limitations can be overcome through partnerships between academics and practitioners (which is how we have conducted our prejudice-reduction work to date; e.g., Green & Wong 2008; Paluck 2006b, 2008; Paluck & Green 2008).

Laboratory research plays an important role in the process of developing and testing interventions, but too often this process stops short of real-world tests. The result is a dearth of rigorously tested interventions and also of rigorously tested theoretical ideas. We urge more research programs in the spirit of psychologists such as Stuart Cook, Kurt Lewin, and Donald Campbell: hypothesis generation through field observation, and intervention testing with parallel laboratory and field experiments. The imperative to test ideas in the

field will keep theories appropriately complex and attuned to real-world conditions, and continually revisiting the laboratory will help to refine understandings of the causal mechanisms at work, which in turn helps inspire new interventions.

In addition to becoming more methodologically rigorous, the study of prejudice reduction must branch out substantively. As our review of the literature demonstrates, the kinds of interventions that have been evaluated do not pit prejudice against its strongest potential adversaries. Studies to date have largely relied on passive and indirect interventions such as coopera-

tive contact. What if interventions were instead to harness forces such as obedience and conformity, the very forces that have been implicated in some of the most notorious expressions of prejudice in world history? If people can be induced to express prejudice at the behest of political leaders, can they also be induced to repudiate prejudice if instructed to do so? If social cues induce conformity to prejudiced norms, can social cues also induce conformity to tolerant norms? The prejudice-reduction literature should be regarded as an opportunity to assess the power and generality of basic psychological theory.

SUMMARY POINTS

1. Notwithstanding the enormous literature on prejudice, psychologists are a long way from demonstrating the most effective ways to reduce prejudice. Due to weaknesses in the internal and external validity of existing research, the literature does not reveal whether, when, and why interventions reduce prejudice in the world.

2. Entire genres of prejudice-reduction interventions, including diversity training, educational programs, and sensitivity training in health and law enforcement professions, have never been evaluated with experimental methods.

3. Nonexperimental research in the field has yielded information about prejudice-reduction program implementation, but it cannot answer the question of what works to reduce prejudice in these real-world settings.

4. Laboratory experiments test a wide range of prejudice-reduction theories and mechanisms with precision. However, researchers should remain skeptical of recommendations based upon environments, interventions, participants, and theories created in laboratory settings until they are supported by research of the same degree of rigor outside of the laboratory.

5. Laboratory research and field research are rarely coordinated; in particular, many prejudice-reduction theories with the strongest support from the laboratory receive scant attention in the field.

6. Field experimentation remains a promising but underutilized approach. Promising avenues for prejudice reduction based on existing field experimentation include cooperative learning, media, and reading interventions.

FUTURE ISSUES

1. More field experimentation can provide evidence that is missing, particularly for the contact hypothesis, peer influence and discussion/dialogue interventions, values and self-worth interventions, social categorization theory, and cognitive training.

2. Theoretical perspectives and more rigorous evaluation methods should be brought to bear on common prejudice interventions such as diversity training; multicultural, antibias, and moral education; sensitivity and cultural competence training; and conflict resolution.

3. Psychologists should look to historical exemplars of theoretically and methodologically rigorous applied prejudice-reduction studies, such as those conducted by Stuart Cook. The hallmark of Cook's work was theoretically grounded randomized field interventions and highly realistic experimental laboratory interventions.

4. In addition to becoming more methodologically rigorous, the study of prejudice reduction must branch out substantively to include more direct interventions based on classic psychological findings (e.g., those that leverage the power of conformity and authority). Researchers should also strive to reduce deeply held prejudices rather than the more transitory prejudices associated with "minimal" groups.

DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Aboud FE, Doyle AB. 1996. Does talk of race foster prejudice or tolerance in children? *Can. J. Behav. Sci.* 28:161–70

Aboud FE, Levy SR. 2000. Interventions to reduce prejudice and discrimination in children and adolescents. In *Reducing Prejudice and Discrimination*, ed. S Oskamp, pp. 269–93. Mahwah, NJ: Erlbaum

Allport G. 1954. *The Nature of Prejudice*. Reading, MA: Addison-Wesley. 537 pp.

Am. Psychol. Assoc. 2001. *Declaration for the UN World Conference Against Racism, Racial Discrimination, Xenophobia, and Related Intolerance*. http://wwwapaorg/pi/oema/racismdeclarationpdf

Aronson E, Blaney N, Stephan W, Sikes J, Snapp M. 1978. *The Jigsaw Classroom*. Thousand Oaks, CA: Sage

Banaji MR. 2001. Implicit attitudes can be measured. In *The Nature of Remembering: Essays in Honor of Robert G. Crowder*, ed. HL Roediger, JS Nairne, I Neath, AM Surprenant, pp. 117–50. Washington, DC: Am. Psychol. Assoc.

Bandura A. 1986. *Social Foundations of Thought and Action*. Englewood Cliffs, NJ: Prentice Hall

Bandura A. 2006. Toward a psychology of human agency. *Perspect. Psychol. Sci.* 1:164–80

Bargal D, Bar H. 1992. A Lewinian approach to intergroup workshops for Arab-Palestinian and Jewish youth. *J. Soc. Issues* 48:139–54

Bargh JA, Chen M, Burrows L. 1996. Automaticity of social behavior: direct effect of trait construct and stereotype activation on action. *J. Personal. Soc. Psychol.* 71:230–44

Batson CD. 1991. *The Altruism Question: Toward a Social-Psychological Answer*. Hillsdale, NJ: Erlbaum. 272 pp.

Baumeister R, Leary M. 1997. Writing narrative literature reviews. *Rev. Gen. Psychol.* 13:311–20

Berry JW. 1984. Multicultural policy in Canada: a social psychological analysis. *Can. J. Behav. Sci.* 16:353–70

Bettencourt A, Brewer M, Croak M, Miller N. 1992. Cooperation and the reduction of intergroup bias: the role of reward structure and social orientation. *J. Exp. Soc. Psychol.* 284:301–19

Bigler RS. 1999. The use of multicultural curricula and materials to counter racism in children. *J. Soc. Issues* 55:687–705

Blair I. 2002. The malleability of automatic stereotypes and prejudice. *Personal. Soc.Psychol. Rev.* 63:242–61

Blanchard FA, Crandall CS, Brigham JC, Vaughn LA. 1994. Condemning and condoning racism: a social context approach to interracial settings. *J. Appl. Psychol.* 79:993–97

Blanchard FA, Lilly T, Vaughn LA. 1991. Reducing the expression of racial prejudice. *Psychol. Sci.* 2:101–5

Bodenhausen GV, Kramer GP, Süsser K. 1994. Happiness and stereotypic thinking in social judgment. *J. Personal. Soc. Psychol.* 66:621–32

Brewer MB, Gaertner SL. 2001. Strategies for reducing intergroup bias. In *Handbook of Social Psychology: Intergroup Processes*, ed. RJ Brown, SL Gaertner, pp. 451–513. Oxford: Blackwell Sci.

Brown RJ, Turner JC. 1979. The criss-cross categorization effect in intergroup discrimination. *Br. J. Soc. Clinical Psychol.* 18:371–83

Cameron L, Rutland A. 2006. Extended contact through story reading in school: reducing children's prejudice towards the disabled. *J. Soc. Issues* 62:469–88

Cameron L, Rutland A, Brown RJ, Douch R. 2006. Changing children's intergroup attitudes towards refugees: testing different models of extended contact. *Child Dev.* 77:1208–19

Clunies-Ross G, O'Meara K. 1989. Changing the attitudes of students towards peers with disabilities. *Aust. Psychol.* 24:273–84

Cook SW. 1971. *The effect of unintended interracial contact upon racial interaction and attitude change*. Proj. No. 5–1320, Final Rep. Washington, DC: U.S. Dep. Health Educ. Welfare

Cook SW. 1978. Interpersonal and attitudinal outcomes in cooperating interracial groups. *J. Res. Dev. Educ.* 12:97–113

Cook SW. 1985. Experimenting on social issues: the case of school desegregation. *Am. Psychol.* 40:452–60

Crandall CS, Eshleman A, O'Brien LT. 2002. Social norms and the expression and suppression of prejudice: the struggle for internalization. *J. Personal. Soc. Psychol.* 82:359–78

Crandall CS, Stangor C. 2005. Conformity and prejudice. In *On the Nature of Prejudice: Fifty Years After Allport*, ed. JF Dovidio, P Glick, LA Rudman, pp. 295–309. Malden, MA: Blackwell Sci.

Crisp RJ, Hewstone M. 1999. Differential evaluation of crossed category groups: patterns processes and reducing intergroup bias. *Group Proc. Intergroup Relat.* 2:303–33

Crisp RJ, Hewstone M. 2007. Multiple social categorization. *Adv. Exp. Soc. Psychol.* 39:163–254

Crosby F, Bromley S, Saxe L. 1980. Recent unobtrusive studies of Black and White discrimination and prejudice: a literature review. *Psychol. Bull.* 87:546–63

Dasgupta N, Greenwald AG. 2001. On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *J. Personal. Soc. Psychol.* 81:800–14

Deutsch M. 1949. A theory of co-operation and competition. *Hum. Relat.* 2:129–52

Devine PG. 1989. Stereotypes and prejudice: their automatic and controlled components. *J. Personal. Soc. Psychol.* 56:5–18

Dobbs M, Crano WD. 2001. Outgroup accountability in the minimal group paradigm: implications for aversive discrimination and social identity theory. *Personal. Soc. Psychol. Bull.* 27:355–64

Dovidio JF, Gaertner SL, Validzic A, Matoka K, Johnson B, Frazier S. 1997. Extending the benefits of recategorization: evaluations, self-disclosure, and helping. *J. Exp. Soc. Psychol.* 33:401–20

Druckman JN, Nelson KR. 2003. Framing and deliberation: how citizens' conversations limit elite influence. *Am. J. Pol. Sci.* 474:729–45

Duncan JG, Boisjoly J, Levy DM, Kremer M, Eccles J. 2003. *Empathy or Antipathy? The Consequences of Racially and Socially Diverse Peers on Attitudes and Behaviors*. Work. Pap., Inst. Policy Res., Northwestern Univ., Chicago, IL

Eagly A, Mladinic A. 1994. Are people prejudiced against women? Some answers from research on attitudes, gender stereotypes, and judgments of competence. In *European Review of Social Psychology*, ed. W Stroebe, M Hewstone, 5:1–35. New York: Wiley

Eisenstadt D, Leippe MR, Rivers JA, Stambush MA. 2003. Counterattitudinal advocacy on a matter of prejudice: effects of distraction commitment and personal importance. *J. Appl. Soc. Psychol.* 33(10):2123–52

Ellis C, Sonnenfield JA. 1994. Diverse approaches to managing diversity. *Hum. Res. Manag.* 33:79–109

Environics Res. Group Ltd. 2001. *Citizenship Advertising Campaign "Canada: We All Belong." PN5048 contract 6C501-010874/001/CB*. Communication Canada

Esses VM, Dovidio JF. 2002. The role of emotions in determining willingness to engage in intergroup contact. *Personal. Soc. Psychol. Bull.* 289:1202–14

Fein S, Spencer SJ. 1997. Prejudice as self-image maintenance: affirming the self through derogating others. *J. Personal. Soc. Psychol.* 73:31–44

Festinger LA. 1957. *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford Univ. Press. 291 pp.

Fisher FL. 1968. Influences of reading and discussion on the attitudes of fifth graders toward American Indians. *J. Educ. Res.* 623:130–34

Flores G, Gee D, Kastner B. 2000. The teaching of cultural issues in US and Canadian medical schools. *Acad. Med.* 75:451–55

Gaertner SL, Dovidio JF. 2000. *Reducing Intergroup Bias: The Common Ingroup Identity Model*. Philadelphia, PA: Psychol. Press. 212 pp.

Gaertner SL, Dovidio JF, Rust MC, Nier J, Banker B, Ward CM. 1999. Reducing intergroup bias: elements of intergroup cooperation. *J. Personal. Soc. Psychol.* 76:388–402

Galinsky AD, Moskowitz GB. 2000. Perspective taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *J. Personal. Soc. Psychol.* 784:708–24

Gardiner GS. 1972. Complexity training and prejudice reduction. *J. Appl. Soc. Psychol.* 2:326–42

Gerber AS, Green DP, Kaplan EH. 2004. The illusion of learning from observational research. In *Problems and Methods in the Study of Politics*, ed. I Shapiro, R Smith, T Massoud, pp. 251–73. New York: Cambridge Univ. Press

Gerrig RJ. 1993. *Experiencing Narrative Worlds*. New Haven, CT: Yale Univ. Press. 294 pp.

Gilliam FD Jr, Iyengar S. 2000. Prime suspects: the impact of local television news on the viewing public. *Am. J. Polit. Sci.* 44:560–73

Gray DB, Ashmore RD. 1975. Comparing the effects of informational role-playing and value-discrepant treatments of racial attitudes. *J. Appl. Soc. Psychol.* 5:262–81

Green DP, Wong JS. 2008. Tolerance and the contact hypothesis: a field experiment. In *The Political Psychology of Democratic Citizenship*, ed. E Borgida. London: Oxford Univ. Press. In press

Greenwald AG, McGhee DE, Schwartz JKL. 1998. Measuring individual differences in implicit cognition: the implicit association test. *J. Personal. Soc. Psychol.* 74:1464–80

Gurin P, Peng T, Lopez G, Nagda BR. 1999. Context identity and intergroup relations. In *Cultural Divides: The Social Psychology of Intergroup Contact*, ed. D Prentice, D Miller, pp. 133–70. New York: Sage

Gwinn CB. 1998. The effect of multicultural literature on the attitudes of second-grade students. Unpubl. PhD dissert., Univ. Minn.

Green MC, Brock TC. 2002. In the mind's eye: transportation-imagery model of narrative persuasion. In *Narrative Impact*, ed. MC Green, JJ Strange, TC Brock, pp. 315–41. Mahwah, NJ: Erlbaum

Hafer C, Begue L. 2005. Experimental research on just-world theory: problems, developments, and future challenges. *Psychol. Bull.* 1311:128–67

Hanover JMB, Cellar DF. 1998. Environmental factors and the effectiveness of workforce diversity training. *Hum. Resour. Dev. Q.* 9:105–24

Hansen F. 2003. Diversity's business case: doesn't add up. *Workforce* 824:28–32

Haring TG, Breen C, Pitts-Conway V, Lee M. 1987. Adolescent peer tutoring and special friend experiences. *J. Assoc. Persons Severe Handicaps* 12:280–86

Hewstone M. 2000. Contact and categorization: social psychological interventions to change intergroup relations. In *Stereotypes and Prejudice: Essential Readings*, ed. C Stangor, pp. 394–418. New York: Psychol. Press

Hornsey M, Hogg M. 2000a. Intergroup similarity and subgroup relations: some implications for assimilation. *Personal. Soc. Psychol. Bull.* 268:948–58

Hornsey M, Hogg M. 2000b. Assimilation and diversity: an integrative model of subgroup relations. *Personal. Soc. Psychol. Rev.* 42:143–56

Houlette M, Gaertner SL, Johnson KS, Banker BS, Riek BM, Dovidio JM. 2004. Developing a more inclusive social identity: an elementary school intervention. *J. Soc. Issues* 601:35–55

Hughes JM, Bigler RS, Levy SR. 2007. Consequences of learning about historical racism among European American and African American children. *Child Dev.* 78:1689–705

Jackman MR. 1994. *The Velvet Glove: Paternalism and Conflict in Gender Class and Race Relations*. Berkeley, CA: Univ. Calif. Press. 425 pp.

Johnson DW, Johnson RT. 1989. *Cooperation and Competition: Theory and Research*. Edina, MN: Interaction Book Co. 265 pp.

Judd CM, Park B, Ryan CS, Brauer M, Kraus S. 1995. Stereotypes and ethnocentrism: diverging interethnic perceptions of African American and White American youth. *J. Personal. Soc. Psychol.* 69:460–81

Karpinski A, Hilton JL. 2001. Attitudes and the Implicit Association Test. *J. Personal. Soc. Psychol.* 81:774–78

Katz P. 2000. Research summary. *Intergroup Relations Among Youth: Summary of a Research Workshop*. New York: Carnegie Corp.

Katz PA, Zalk SR. 1978. Modification of children's racial attitudes. *Dev. Psychol.* 145:447–61

Kawakami K, Dovidio JF, Moll J, Hermsen S, Russin A. 2000a. Just say no to stereotyping: effects of training in trait negation on stereotype activation. *J. Personal. Soc. Psychol.* 78:871–88

Kawakami K, Phills CE, Steele JR, Dovidio JF. 2000b. Close distance makes the heart grow fonder: improving implicit racial attitudes and interracial interactions through approach behaviors. *J. Personal. Soc. Psychol.* 926:957–71

Kelman HC, Fisher RJ. 2003. Conflict analysis and resolution. In *Political Psychology*, ed. DO Sears, L Huddy, R Jervis, pp. 315–56. London: Oxford Univ. Press

Kiselica MS, Maben P. 1999. Do multicultural education and diversity training appreciation training reduce prejudice among counseling trainees? *J. Ment. Health Couns.* 213–40

Kuklinski JH, Hurley NL. 1996. It's a matter of interpretation. In *Political Persuasion and Attitude Change*, ed. D Mutz, P Sniderman, R Brody, pp. 125–44. Ann Arbor: Mich. Univ. Press

Kulik CT, Perry EL, Bourhis AC. 2000. Ironic evaluation processes: effects of thought suppression on evaluations of older job applicants. *J. Org. Behav.* 21:689–711

Landis D, Day HR, McGrew PL, Thomas JA, Miller AB. 1976. Can a black "culture assimilator" increase racial understanding? *J. Soc. Issues* 32:169–83

Levy SR. 1999. Reducing prejudice: lessons from social-cognitive factors underlying perceiver differences in prejudice. *J. Soc. Issues* 55:745–65

Levy SR, Stroessner S, Dweck CS. 1998. Stereotype formation and endorsement: the role of implicit theories. *J. Personal. Soc. Psychol.* 74:421–36

Levy SR, West TL, Ramirez LF, Pachankis JE. 2004. Racial and ethnic prejudice among children. In *The Psychology of Prejudice and Discrimination: Racism in America*, Vol. 1, ed. J Chin, pp. 37–60. Westport, CT: Praeger Publ.

Liebkind K, McAlister AL. 1999. Extended contact through peer modelling to promote tolerance in Finland. *Eur. J. Soc. Psychol.* 295–6:765–80

Lustig I. 2003. *The influence of studying foreign conflicts on students' perceptions of the Israeli Palestinian conflict*. Unpubl. master's thesis: Univ. Haifa, Israel

Maass A. 1999. Linguistic intergroup bias: stereotype perpetuation through language. *Adv. Exp. Soc. Psychol.* 31:79–121

Marcus-Newhall A, Miller N, Holtz R, Brewer MB. 1993. Cross-cutting category membership with role assignment: a means of reducing intergroup bias. *Br. J. Soc. Psychol.* 32:125–46

Mays L, Henderson EH, Seidman SK, Steiner VJ. 1975. *An Evaluation Report on Vegetable Soup: The Effects of a Multi-Ethnic Children's Television Series on Intergroup Attitudes of Children*. Dept. Educ., Albany, NY

McAlister A, Ama E, Barroso C, Peters RJ, Kelder S. 2000. Promoting tolerance and moral engagement through peer modeling. *Cultur. Divers. Ethnic Minor. Psychol.* 64:363–73

McCauley C, Wright M, Harris M. 2000. Diversity workshops on campus: a survey of current practice at U.S. colleges and universities. *Coll. Student J.* 34:100–14

Mendelberg T. 2001. *The Race Card*. Princeton, NJ: Princeton Univ. Press

Miller N, Brewer MB. 1986. Categorization effects on ingroup and outgroup perception. In *Prejudice Discrimination and Racism*, ed. JF Dovidio, SL Gaertner, pp. 209–230. Orlando, FL: Academic

Miller N, Harrington HJ. 1990. A model of social category salience for intergroup relations: empirical tests of relevant variables. In *European Perspectives in Psychology*, ed. R Takens, pp 205–20. New York: Wiley

Monteith MJ, Deneen NE, Tooman G. 1996. The effect of social norm activation on the expression on opinions concerning gay men and blacks. *Basic Appl. Soc. Psychol.* 18:267–88

Monteith MJ, Sherman J, Devine PG. 1998. Suppression as a stereotype control strategy. *Personal. Soc. Psychol. Rev.* 2:3–82

Monteith MJ, Zuwerink JR, Devine PG. 1994. Prejudice and prejudice reduction: classic challenges contemporary approaches. In *Social Cognition: Impact on Social Psychology*, ed. PG Devine, DL Hamilton, pp. 323–46. San Diego, CA: Academic

Morris L, Romero J, Tan DL. 1996. Changes in attitude after diversity training. *Training Dev.* 509:54–55

Mullen B, Migdal M, Hewstone M. 2001. Crossed categorization versus simple categorization and intergroup evaluations: a meta-analysis. *Eur. J. Soc. Psychol.* 31:721–36

Nagda B, Kim C, Truelove Y. 2004. Learning about difference: learning with others, learning to transgress. *J. Soc. Issues* 60:195–214

Nagda B, Zúñiga X. 2003. Fostering meaningful racial engagement through intergroup dialogues. *Group Process Intergr.* 6:111–28

Nier JA, Gaertner SL, Dovidio JF, Banker BS, Ward CM, Rust MC. 2001. Changing interracial evaluations and behavior: the effects of a common group identity. *Group Process Interg.* 44:299–316

Olson MA, Fazio RH. 2008. Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personal. Soc. Psychol. Bull.* 32:421–33

Oskamp S. 2000. *Reducing Prejudice and Discrimination*. Mahwah, NJ: Erlbaum. 370 pp.

Pagtolun-An IM, Clair JM. 1986. An experimental study of attitudes toward homosexuality. *Dev. Behav.* 7:121–35

Paluck EL. 2006a. Diversity training and intergroup contact: a call to action research. *J. Soc. Issues* 623:439–51

Paluck EL. 2006b. *Peer pressure against prejudice: a field experimental test of a national high school prejudice reduction program*. Work. Pap., Harvard Univ., Cambridge, MA

Paluck EL. 2008. Reducing intergroup prejudice and conflict using the media: a field experiment in Rwanda. *J. Personal. Soc. Psychol.* In press

Paluck EL, Green DP. 2008. *Deference dissent and dispute resolution: a field experiment on a mass media intervention in Rwanda*. Work. Pap., Harvard Univ., Cambridge, MA

Pedersen A, Walker I, Wise M. 2005. "Talk does not cook rice": beyond antiracism rhetoric to strategies for social action. *Aust. Psychol.* 401:20–30

Pettigrew TF, Tropp LR. 2006. A meta-analytic test of intergroup contact theory. *J. Personal. Soc. Psychol.* 90:751–83

Powers DA, Ellison CG. 1995. Interracial contact and black racial attitudes: the contact hypothesis and selectivity bias. *Soc. Forces* 74:205–26

Price EG, Beach MC, Gary TL, Robinson KA, Gozu A, Palacio A. 2005. A systematic review of the methodological rigor of studies evaluating cultural competence training of health professionals. *Acad. Med.* 806:579–86

Puma MJ, Jones CC, Rock D, Fernandez R. 1993. Prospects: The congressionally mandated study of educational growth and opportunity. *Interim rep.*, Bethesda, MD: Abt Assoc.

Rich Y, Kedem P, Shlesinger A. 1995. Enhancing intergroup relations among children: a field test of the Miller-Brewer model. *Int. J. Intercult. Relat.* 194:539–53

Roberts K. 2000. *Toward Competent Communities: Best Practices for Producing Community-Wide Study Circles*. Lexington, KY: Topsfield Found.

Rokeach M. 1971. Long-range experimental modification of values attitudes and behavior. *Am. Psychol.* 26:453–59

Rokeach M. 1973. *The Nature of Human Values*. New York: Free Press

Roseth C, Johnson D, Johnson R. 2008. Promoting early adolescents' achievement and peer relationships: the effects of cooperative competitive and individualistic goal structures. *Psychol. Bull.* 1342:223–46

Rothbart M, John OP. 1985. Social categorization and behavioral episodes: a cognitive analysis of the effects of intergroup contact. *J. Soc. Issues* 413:81–104

Rudman LA, Ashmore RD, Gary ML. 2001. "Unlearning" automatic biases: the malleability of implicit prejudice and stereotypes. *J. Personal. Soc. Psychol.* 815:856–68

Ruiz-Belda MA, Fernandez-Dols JM, Carrera P, Barchard K. 2003. Spontaneous facial expressions of happy bowlers and soccer fans. *Cogn. Emot.* 17:315–26

Schaller M, Asp CH, Rosell MC, Heim SJ. 1996. Training in statistical reasoning inhibits the formation of erroneous group stereotypes. *Personal. Soc. Psychol. Bull.* 22:829–44

Sears DO. 1986. College sophomores in the laboratory: influences of a narrow data-base on social psychology's view of human nature. *J. Personal. Soc. Psychol.* 51:515–30

Sears DO. 1988. Symbolic racism. In *Handbook of Social Psychology*, ed. G. Lindzey, E Aronson, 5:315–458. Reading, MA: Addison-Wesley. Rev. ed.

Selltiz C, Cook S. 1948. Can research in social science be both socially meaningful and scientifically useful? *Am. Soc. Rev.* 13:454–59

Shadish WR, Cook TD, Campbell DT. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton-Mifflin. 623 pp.

Sherif M, Harvey OJ, White BJ, Hood WR, Sherif CW. 1961. *Intergroup Conflict and Cooperation: The Robbers Cave Experiment*. Norman: Univ. Oklahoma Book Exchange. 263 pp.

Sidanius J, Pratto F. 1999. *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression*. New York: Cambridge Univ. Press. 416 pp.

Sinclair L, Kunda Z. 1999. Reactions to a Black professional: motivated inhibition and activation of conflicting stereotypes. *J. Personal. Soc. Psychol.* 77:885–904

Slavin RE, Leavey MB, Madden NA. 1984. Combining cooperative learning and individualized instruction: effects on student mathematics achievement, attitudes, and behaviors. *Element. Sch. J.* 84:409–22

Slone M, Tarrasch R, Hallis D. 2000. Ethnic stereotypic attitudes among Israeli children: two intervention programs. *Merrill-Palmer Q.* 46:370–89

Smith E. 1993. Social identity and social emotions: towards new conceptualizations of prejudice. In *Affect, Cognition, and Stereotyping*, ed. D Mackie, D Hamilton, pp. 297–315. New York: Academic

Sniderman PM, Piazza T. 1993. *The Scar of Race*. Cambridge, MA: Harvard Univ. Press

Son Hing LS, Li W, Zanna MP. 2002. Inducing hypocrisy to reduce prejudicial responses among aversive racists. *J. Exp. Soc. Psychol.* 381:71–78

Stangor C, Sechrist GB, Jost JT. 2001. Changing racial beliefs by providing consensus information. *Personal. Soc. Psychol. Bull.* 27:484–94

Steele CM. 1998. The psychology of self-affirmation: sustaining the integrity of the self. In *Advances in Experimental Social Psychology, Vol. 21: Social Psychological Studies of the Self: Perspectives and Programs*, ed L Berkowitz, pp. 261–302. San Diego, CA: Academic

Stephan WG. 1999. *Reducing Prejudice and Stereotyping in Schools*. New York: Teachers Coll. Press. 142 pp.

Stephan WG, Finlay K. 1999. The role of empathy in improving intergroup relations. *J. Soc. Issues* 554:729–44

Stephan WG, Stephan CW. 1984. The role of ignorance in intergroup relations. In *Groups in Contact: The Psychology of Desegregation*, ed. N Miller, MB Brewer, pp. 229–57. Orlando, FL: Academic

Stephan WG, Stephan CW. 2001. *Improving Intergroup Relations*. Thousand Oaks, CA: Sage. 360 pp.

Sternberg RJ. 2003. A duplex theory of hate: development and application to terrorism massacres and genocide. *Rev. Gen. Psychol.* 73:299–328

Stevenson MR. 1988. Promoting tolerance for homosexuality: an evaluation of intervention strategies. *J. Sex Res.* 254:500–11

Stewart TL, LaDuke JR, Bracht C, Sweet BAM, Gamarel KE. 2003. Do the "eyes" have it? A program evaluation of Jane Elliot's "Blue Eyes/Brown Eyes" diversity training exercise. *J. Appl. Soc. Psychol.* 339:1898–921

Strange JL. 2002. How fictional tales wag real-world beliefs. In *Narrative Impact: Social and Cognitive Foundations*, ed. MC Green, J Strange, T Brock, pp. 263–86. Mahwah, NJ: Erlbaum

Tajfel H. 1970. Experiments in intergroup discrimination. *Sci. Am.* 223:96–102

Towles-Schwen T, Fazio RH. 2006. Automatically activated racial attitudes as predictors of the success of interracial roommate relationships. *J. Exp. Soc. Psychol.* 42:698–705

van Wieringen JCM, Schulpen TWJ, Kuyvenhoven MM. 2001. Intercultural training of medical students. *Med. Teach.* 23:80–82

Vescio TK, Judd CM, Kwan VSY. 2004. Categorization and intergroup bias in crossed contexts: evidence of reductions in the strength of categorization but not intergroup bias. *J. Exp. Soc. Psychol.* 40:478–96

Vescio TK, Sechrist GB, Paolucci MP. 2003. Perspective taking and prejudice reduction: the mediational role of empathy arousal and situational attributions European. *J. Soc. Psychol.* 33 455–72

Warring DW, Johnson D, Maruyama G, Johnson R. 1985. Impact of different types of cooperative learning on cross-ethnic and cross-sex relationships. *J. Educ. Psychol.* 77:53–59

Wham MA, Barnhart J, Cook G. 1996. Enhancing multicultural awareness through the storybook reading experience. *J. Res. Dev. Educ.* 301:1–9

Wikfors EL. 1998. Effects of a disabilities awareness workshop: employees' knowledge attitude and interaction. PhD dissert., Univ. Florida. *Dissert. Abstracts Int.* 58:2495

Wilder DA. 1986. Social categorization: implications for creation and reduction of intergroup bias. In *Advances in Experimental Social Psychology, Vol. 19*, ed. L Berkowitz, pp. 239–355. New York: Academic

Wittenbrink B, Judd CM, Park B. 2001. Spontaneous prejudice in context: variability in automatically activated attitudes. *J. Personal. Soc. Psychol.* 81:815–27

Wright SC, Aron A, McLaughlin-Volpe T, Ropp SA. 1997. The extended contact effect: knowledge of cross-group friendships and prejudice. *J. Personal. Soc. Psychol.* 73:73–90

Yawkey TD. 1973. Attitudes toward Black Americans held by rural and urban white early childhood subjects based upon multi-ethnic social studies materials. *J. Negro Educ.* 42:164–69

Zillmann D. 1991. Empathy: affect from bearing witness to the emotions of others. In *Responding to the Screen: Reception and Reaction Processes*, ed. J Bryant, D Zillmann, pp. 335–67. Hillsdale, NJ: Erlbaum