Causal effect of intergroup contact on exclusionary attitudes

Ryan D. Enos[1]

Department of Government, Harvard University, Cambridge, MA 02138

Edited by David D. Laitin, Stanford University, Stanford University, CA, and approved January 21, 2014 (received for review September 18, 2013)

The effect of intergroup contact has long been a question central to social scientists. As political and technological changes bring increased international migration, understanding intergroup contact is increasingly important to scientific and policy debates. Unfortunately, limitations in causal inference using observational data and the practical inability to experimentally manipulate demographic diversity has limited scholars' ability to address the effects of intergroup contact. Here, I report the results of a randomized controlled trial testing the causal effects of repeated intergroup contact, in which Spanish-speaking confederates were randomly assigned to be inserted, for a period of days, into the daily routines of unknowing Anglo-whites living in homogeneous communities in the United States, thus simulating the conditions of demographic change. The result of this experiment is a significant shift toward exclusionary attitudes among treated subjects. This experiment demonstrates that even very minor demographic change causes strong exclusionary reactions. Developed nations and politically liberal subnational units are expected to experience a politically conservative shift as international migration brings increased intergroup contact.

intergroup attitudes | field experiment | political science | psychology | immigration

Under general conditions, contact between members of different social-identity groups is positively correlated with between-group discriminatory attitudes and behaviors (1) and is often related to reduced social capital (2), inefficient resource distribution (3), lack of democratic consensus (4), and violent conflict (5, 6).

Intergroup contact is a common outcome of demographic diversity, which has long been a central social issue in plural societies (7) and a political issue since the rise of nation states (8). Furthermore, demographic change, and the accompanying intergroup contact, is an increasingly important issue as economic and technological changes spur population movements from less economically to more economically advanced nations.

However, the problem of identifying the causal effects of intergroup contact is among the most vexing in social science. Despite its widely accepted importance in both academic and policy circles, this problem has remained intractable—even with the many advances in social science—because of an inability to experiment on intergroup contact (9). In addition to this inability, the causal mechanisms remain disputed: Proposed mechanisms for the correlation between intergroup contact and intergroup conflict include evolutionary (10, 11), cognitive (12), economic (13, 14), and political (15) processes.

Here I report the results of a randomized controlled trial identifying the effects of increased intergroup contact under conditions similar to demographic change. The experiment randomly assigned repeated real-world contact, over an extended period, between individuals of different ethnic groups in the United States. People exposed to the outgroup underwent a strong exclusionary shift, relative to a control group, in their attitudes toward the outgroup. However, this effect may have weakened with repeat contact. Furthermore, the subjects exposed to the outgroup did not similarly alter their opinions about other groups that were unrelated to the treatment. The results identify a causal effect of intergroup contact under real-world conditions and addresses the reactions that should be expected of native populations exposed to increased cross-national immigration.

Despite a large body of literature and important findings across a number of fields, observational studies of the effects of diversity can suffer from problems of selection, making it very difficult to understand the causal effects of diversity on behavior (16–19). Complicating matters further, some theories have even proposed that under specific conditions, intergroup contact may lead to more positive attitudes toward outgroup members (20, 21).

Even when clear observational correlations can be drawn between demographic change and behavior, we still cannot know if it is the contact itself that causes the behavior. For example, the widely observed antagonistic behavior toward African Americans by city-dwelling whites in the United States in the 1960s (22–24) may not have been directly caused by demographic change. Another plausible explanation is that opportunistic politicians exploited potential intergroup hostilities (25, 26), thereby causing hostility that would not otherwise have existed. Some scholars have made use of longitudinal studies of attitudes (27), but even these, although yielding valuable insights, cannot overcome problems of selection (19) and other confounding issues (17).

Scholars have recognized this difficulty and have extensively used laboratory experiments to study the phenomenon of intergroup contact (28). These demonstrations have greatly enhanced the theoretical understanding of intergroup conflict by detecting short-term correlations between group-based identities and individual discriminatory behavior, and by refining our understanding of the conditions under which conflict occurs (29, 30). However, these experiments lack the important externally valid condition of repeated, interpersonal contact that accompanies demographic change (9, 21). This condition is important because real-world demographic change involves the extended

Significance

There is generally conflict when members of different social groups, such as racial, ethnic, or religious groups, come in contact in the same geographic area. This phenomenon is commonly observed across a variety of settings. However, the cause of such conflict is poorly understood: Some theorists have argued that contact between groups is insufficient to cause conflict and that, under certain conditions contact may lead to improved intergroup relations. Although most theories of contact propose that repeated contact between individuals is important to the disposition of intergroup attitudes, experimenting on the effects of repeated contact has proven difficult. Here, I report a randomized controlled trial that assigns repeated intergroup contact between members of different ethnic groups. The contact results in exclusionary attitudes toward the outgroup.

Author contributions: R.D.E. designed research, performed research, analyzed data, and wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission.

[1]E-mail: renos@gov.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1317670111/-/DCSupplemental.

interaction—or potential for interaction—between social groups, even if not between the same individuals. This extended interaction, under the right conditions, may lead to a reduction in prejudicial attitudes because of stereotype reduction (21, 31) or simply because a reduction in the novelty of contact reduces the salience of the outgroup (32). Scholars have also raised concerns about the generality of laboratory experiments for many forms of intergroup conflict (33, 34).

Recognizing this limitation, scholars have resourcefully made use of "natural experiments," where the variation in contact can be seen "as if random" and the contact between group members is repeated. These experiments have included such circumstances as the "as if random" assignment of college roommates of different races (35–37). Although such studies are valuable, their limitations include the external validity of the contact: the narrowly applicable situation of cohabitation rather than the more common experience of living in a diverse community (34), the unrepresentative nature of United States college students (38), and the not truly random nature of the assignment (39).

A widely studied mechanism for the connection between diversity and intergroup conflict, recognized by economists (29), political scientists (40), psychologists (41), and sociologists (42), is that the mere presence of a proximate outgroup leads to antipathy. This mechanism is often referred to as "Group Threat." With this mechanism, elite manipulation, economic or political pressure, or other such social and political factors are not necessary for conflict. Rather, the mere presence of the outgroup causes negative reactions, possibly because proximity increases the salience of the outgroup, thereby activating negative stereotypes (43, 44).

Ideally, to experimentally test for the effects of intergroup contact as experienced in the real world and to test for Group Threat or other mechanisms, a researcher would randomly assign some people to experience demographic change and then observe the subsequent behavioral changes in those people compared with a control group; for example, perhaps by randomly acquiring homes in demographically homogeneous communities and moving in members of a demographic outgroup. However, practical and ethical concerns make such an experiment infeasible.

In the experiment reported here, I approximate this hypothetical experiment by randomly assigning some people to repeatedly encounter members of a demographic outgroup, thereby simulating the effects of demographic change. This was accomplished by sending a small number of Spanish-speaking confederates to commuter train stations in homogeneously Anglo communities every day, at the same time, for 2 wk.

The experiment leveraged the tendency for commuters to ride the same train every day. I treated certain trains by assigning pairs of Spanish-speaking persons to visit the same train stations at the same time every day. Within each train station, these experimental confederates were the same persons every day. Other trains were randomly assigned to the control condition and had no intervention at the stations. Subjects were surveyed about their socio-political attitudes before and after the treatment. With this design, subjects were exposed to the same Spanish-speaking persons in a location near their homes for an extended period, as would be the situation if immigrants had moved into their neighborhood and used the public transportation. With this design, I experimentally manipulated the conditions of demographic change and, by comparing changes in survey responses before and after the treatments, I identified the effect of exposure to these Spanish-speaking persons.

The experiment was conducted in the Boston, MA metropolitan area in homogeneously Anglo communities. The growing Latino community in the United States is bringing demographic change (45), but the change is uneven, with some communities relatively unaffected. The relatively stable homogeneity of the chosen area allows for experimentation. The population of Massachusetts was 9.9% Hispanic origin in 2011, compared with 16.7% of the population of the United States as a whole. The Boston suburban communities with commuter rail stations were, on average, just 4.4% Hispanic in 2010 (Fig. 1). The skewed distribution of Latinos in the United States (Fig. 1) shows the relevance of this test for understanding the impact of demographic change, both in the United States and in other countries, with influxes of immigration. Although the mean Census Tract in the United States is 13.8% Hispanic, the median is only 5.2%. The Census Tracts used in this experiment had a mean of just 2.8% Hispanic, making the communities tested here both demographically typical and representative of the type of community in which demographic change has not already occurred. The observed response to simulated demographic change in still relatively homogeneous communities can shed light on whether homogeneous localities will experience changes in attitudes toward immigrants as population change occurs.

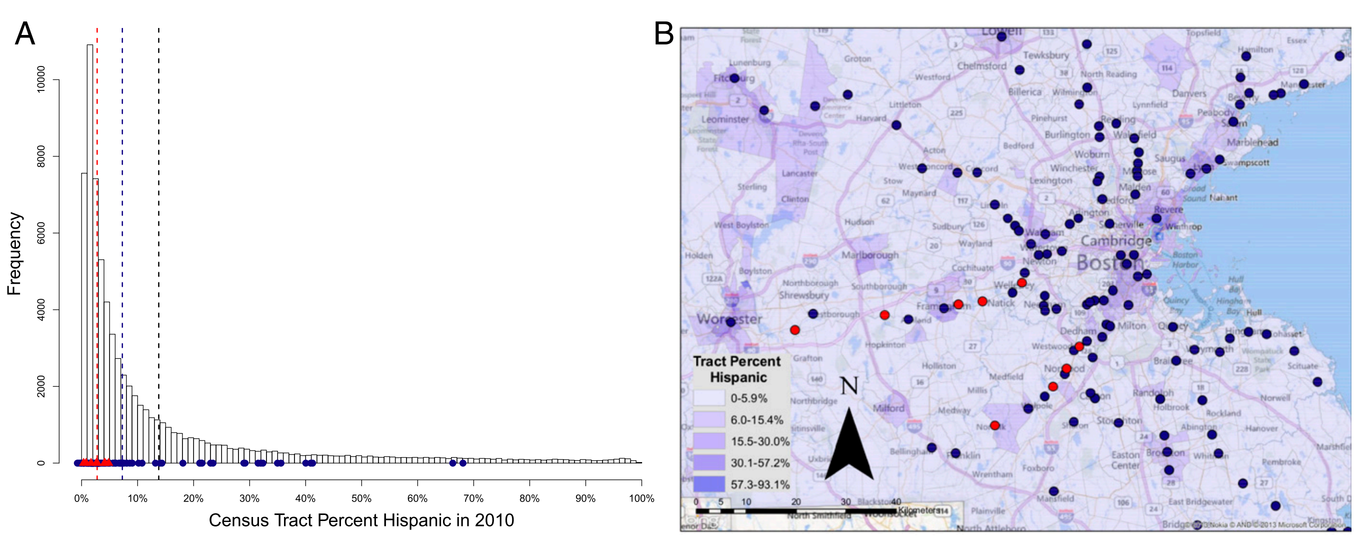

Fig. 1. Distribution of Hispanic persons (of any race) by Census Tract in the United States and in the experiment location in 2010. (*A*) Frequency of Census Tracts by percent Hispanic in the United States. Purple dots are percent Hispanic in Census Tracts containing an MBTA station and red triangles are the stations used in the experiment. Dotted vertical lines indicate the means in the United States (black), all tracts containing MBTA stations (purple), and in stations used in the experiments (red). (*B*) Hispanic persons (of any race) by Census Tract and MBTA stations used in experiment area. Purple dots are MBTA stations and red dots are the stations used in the experiment.

Results

Experiment. Nine commuter rail stations were selected for the experiment (Fig. 1). In each station, several trains come through during the morning rush hour. Each of these trains is a potential "treatment unit" and randomization occurs at the treatment-unit level. Within each station, I examined the potential treatment units and selected two trains, so that there was a matched pair of units within each station. Under the assumption that people with similar characteristics tend to ride the train at the same time, I selected pairs that were close together in time so that the treatment units within each station would have similar passengers. Within a matched pair of train times at each station, one was randomly assigned to treatment and one to control, resulting in 18 matched pairs of train times. This design means that we should expect subjects in the treatment and control conditions to be, in expectation, identical. Balance between treatment and control is shown in *Materials and Methods*.

The passengers in the experimental sample were self-reported 83% white and 4% Hispanic. Survey results confirmed that routinized behavior is common among them. Pretreatment, 88% said they took the train every weekday and 98% said they took it at least three times a week. Posttreatment, respondents indicated that over the 10 working days of the experiment, 78% had caught the train at the exact same time every day, and 96% indicated they had missed their usual train two or fewer times.

I hired pairs of native Spanish-speaking confederates to wait on the platform with the commuters assigned to treatment. These confederates were blind to the hypothesis and purpose of the experiment. (After the experiment, the confederates were fully debriefed on the purpose of the experiment. They were compensated at an hourly rate for their time and paid a bonus for completing the entire task successfully.) The confederates successfully treated every assigned unit on every day of the experiment. They were given no specific instructions about speaking or otherwise interacting with anyone on the platform. They did report conversations that occasionally occurred when other passengers asked for directions or other such normal interactions that might occur at a train station.

A crucial feature of this experimental design is that people on the platform assume that the confederates are Hispanic. In the *Supporting Information*, I establish that the confederates were likely to have been seen as Hispanic foreigners by Anglos at the train stations, but were not extraordinary-looking persons who would have been unusually threatening compared with similar Anglo or Hispanic whites.

Five days before the beginning of the treatment, subjects on the train platform, in both the treatment and control groups, were induced by payment to complete a Web-based survey (T1). After the treatment, subjects who completed the survey and provided a valid e-mail address were then invited, via e-mail, to complete a second round of the survey, with the same attitudinal questions (T2). Among the subjects eligible to take the second round, half were randomly assigned to receive the second survey after 3 d of treatment and half were assigned to receive the second survey after 2 wk (10 working days) of treatment.

With the survey, I collected pretreatment political and demographic characteristics, opinions about their community, and posttreatment questions about commuting during the period of the treatment. The survey also collected three dependent variables about exclusionary policies:

i) "Do you think the number of immigrants from Mexico who are permitted to come to the United States to live should be increased, left the same, or decreased?"
ii) "Would you favor allowing persons that have immigrated to the United States illegally to remain in the country if they are employed and have no criminal history?"
iii) "Some people favor a state law declaring English as the Official Language. Some other people oppose such a law. Would you favor such a law?"

No other variables about immigration policy or exclusionary attitudes were collected.

Results. The experiment shifted the attitudes of the treatment group relative to the control in an exclusionary direction between T1 and T2 on all of the policy questions and especially strongly for the first two questions. The results are presented in the first "All respondents" column of Table 1. This column lists the average treatment effect (ATE) with the *P* value of the estimate in parentheses. Positive numbers represent more exclusionary attitudes. The T1 level of the responses with SDs is listed in the second "All respondents" column of Table 1. The ATEs represent changes of 0.330, 0.201, and 0.082 in normalized SD units.

Treated subjects were far more likely to advocate a reduction in immigration from Mexico and were far less likely to indicate that illegal immigrants should be allowed to remain in this country. The ATEs and associated SEs allow me to reject the Null Hypothesis of no effect with a high degree of confidence. The ATE on favoring English as an official language, although in the same exclusionary direction, is smaller and does not allow me to reject the Null Hypothesis. However, baseline rates for this question are considerably higher (0.610, 0–1 scale) than for the other questions, indicating relatively high support for English as an official language, regardless of treatment.

The confederates reported, without directly being asked, that persons noticed and displayed some unease with them: for example reporting that "Because we are chatting in Spanish, they look at us. I don't think it is common to hear people speaking in Spanish on this route." After the experiment, the confederates reported that other passengers were generally friendly to them but also reported that they felt people noticed them for "not being like them and being Latino."

Separately, I limited the analysis to people who indicated that they wait on the platform while waiting for the train rather than waiting in their cars in nearby parking lots, because people who remain in their cars are less likely to be exposed to the treatment

POLITICAL SCIENCES

Table 1. Experiment results

Question	All respondents	Waits on platform	All respondents
Question	ATE (P)*	CATE (P)	T1 levels (SD)
Number of immigrants be increased?†	0.09 (0.008)	0.083 (0.012)	0.489 (0.272)
Undocumented immigrants allowed to stay?	0.073 (0.016)	0.098 (0.016)	0.441 (0.362)
English as official language?	0.03 (0.27)	0.043 (0.152)	0.619 (0.364)
n	109	100	109

In the first "All respondents" column, ATE represents responses in T2-T1 for the treatment group compared with the control group for the entire experimental sample. Positive values mean a more politically conservative response. In the "Waits on platform" column, CATEs are the Conditional Average Treatment Effects for persons who said they stand on the platform, rather than wait in their cars. In the second "All respondents" column, T1 levels and SDs for each variable for all respondents. All variables scaled 0–1.

**P* values from a one-tailed test against the Null Hypothesis of no effect are in parentheses.

†Each of the questions allowed responses on a five-point scale ranging from strongly agree to strongly disagree (exact answers were changed to be appropriate to the actual question).

as intended. As expected in this subset, the results, reported in the "Waits on platform" column of Table 1 are slightly stronger.

Time Effects. Subjects were randomly assigned to be surveyed when their exposure to the treatment had either been 3 d or 10 d. Comparing these responses can provide some insight into the longer-term effects of this intergroup contact. I display these results in Fig. 2. The 3-d treatment is represented by solid circles and the 10-d treatment by open circles. The lines represent the 95%-confidence interval of the ATE.

Although the 10-d treatment still moves opinions in an exclusionary direction, the effects are considerably stronger after 3 d than after 10 d, perhaps indicating that repeated exposure to an outgroup can mitigate initial negative reactions. The reduced sample sizes mean that inference should be made with caution and these results are only suggestive; only for the question about children of undocumented immigrants is the associated P value marginally significant ($P = 0.094$, two-tailed test for the Null Hypothesis of no difference in effects between treatments). However, these groups were assigned randomly, meaning the effect should be considered the result of the difference in length of exposure to the treatment.

The responses collected from the confederates provide additional evidence that survey respondents were becoming comfortable with their presence. In the waning days of the experiment, confederates commonly reported observations, such as "people have started to recognize and smile to us," and that a passenger who spoke to them said that "the longer you see the same person every day the more confident you feel to greet and say 'hi' to them." These interactions suggest that the initial aversion had diminished.

Discussion

This experiment demonstrated that exclusionary attitudes can be stimulated by even very minor, noninvasive demographic change: in this case, the introduction of only two persons. Overtly threatening behavior by newcomers is not a necessary component for the stimulation of exclusionary attitudes. By examining subsets of the data in the *Supporting Information*, I also demonstrate that political or economic competition do not appear to be necessary conditions for the stimulation of exclusionary attitudes. Rather, it seems that—consistent with theories of Group Threat—the casual presence of outsiders causes an exclusionary reaction, perhaps because of the activation of negative stereotypes. Further research using this experimental paradigm may help to establish the mechanism underlying this relationship.

Perhaps the introduction of an unfamiliar person, regardless of ethnicity, would shift attitudes in a politically conservative direction. The counterfactual implication of this hypothesis is that, had I sent Anglo whites to the train stations, respondents would have reacted in the same way. There are, at least, two reasons that this counter hypothesis is unlikely.

First, if this counter hypothesis were true, we should expect to see a shift in other attitudes as well. This is because it is unlikely that introducing Anglos would directly cause a backlash against immigrants from Mexico. Rather, the introduction of an outsider would have to cause a general shift toward political conservatism that would also shift attitudes about immigration. If it is a general stimulating of conservative attitudes causing the shift toward exclusion, then nonimmigrant-related attitudes should also shift in a conservative direction. There is little evidence that the treatment caused a general conservative shift in attitudes. I asked respondents in both T1 and T2 about their political ideology, party identification, potential vote choice in the 2012 presidential election, and their perceived threat from Asian Americans, African Americans, and persons of Middle Eastern ancestry. These variables do not show a similar change between T1 and T2, indicating that although exclusionary attitudes were stimulated, other politically conservative attitudes were not.

Second, the hypothesis is also unlikely when considering the regular flow of persons to and from these communities. American Community Survey data indicate that the communities in this experiment each experience the in-migration of hundreds of non-Hispanic people every year (compared with fewer than 10 Hispanics). If even a small portion of these people use the train, then it is unlikely that the counterfactual introduction of two additional non-Hispanic train riders would be novel enough to induce an exclusionary response.

Intergroup conflict is correlated with demographic heterogeneity. However, the difficulty in identifying the causal effects of diversity means that scientists have been limited in their ability to answer questions about the likely sociopolitical effects of increasing diversity in national and subnational units in Western democracies. For example, will relatively racially homogeneous states in the United States, such as Massachusetts, adopt the exclusionary policies of states with large Latino populations, such as Arizona? The findings here indicate that continued demographic change in Western nations will be accompanied by impulses for intergroup exclusion and that regions predicted to become more diverse should expect initial conflict. However, these results also suggest that more prolonged contact or interpersonal interaction can diminish initial exclusionary impulse. This last fact is important for policy makers as they consider the public policy with respect to immigrant incorporation. Given a goal of intergroup harmony, further exploration should be given to public policies that encourage interpersonal contact and incorporation.

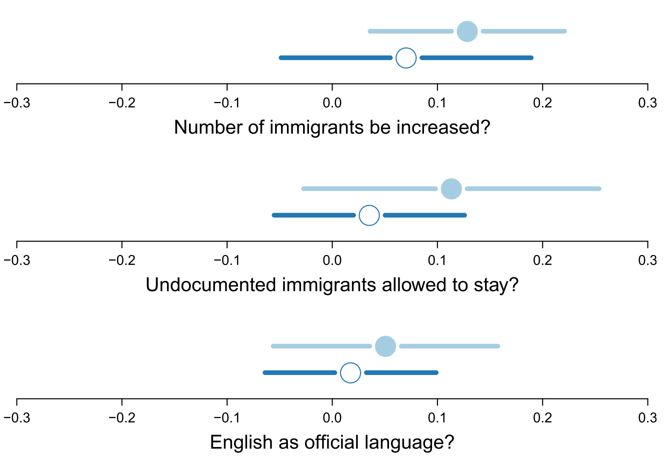

Fig. 2. Time effects. ATE and 95% confidence intervals for 3-d treatment (solid circle) and 10-d treatment (open circle). P values from top to bottom generated from a two-tailed test against the Null Hypothesis of no difference in effect between the 3-d and 10-d treatments are $P = 0.195$, 0.094, and 0.305. n = 55 for 3-d dose and 54 for 10-d dose. Confidence intervals are constructed by drawing the 2.5% and 97.5% quantiles from the randomization distribution.

Materials and Methods

Location and Confederates. Written informed consent was obtained from all subjects. Before beginning the experiment, the design was approved by Harvard University Committee on the Use of Human Subjects. The experiment was conducted in July and August 2012 at stations belonging to the Massachusetts Bay Transit Authority (MBTA), which extends out of Boston on 12 lines with 134 stations (Fig. 1). These stations vary in size, but those in the experimental sample were in relatively low-density neighborhoods and had only one or two platforms. Many stations have large parking lots surrounding the stations to allow commuters to "park-n-ride" to Boston. The number of riders boarding any particular train varies significantly across time and station.

When researchers visited all stations before commencing treatment, they observed anywhere between 10 to over 100 riders waiting. The impression of the researchers, reinforced by the results of the survey, is that riders at these stations tend to be familiar with each other, either recognizing each other by sight or sometimes having regular conversational partners. Confederates repeatedly reported observations, such as "people in the train know each other very well and they are close friends; they get on the train and greet each other as if they were friends for life."

Table 2. Covariate balance across treatment conditions

Condition	Control	Treatment	Standard difference*	*Z* score
Liberal[†]	0.47	0.47	0.01	0.03
Republican	0.17	0.19	0.05	0.24
Obama disapprove	0.27	0.29	0.05	0.24
Ride MBTA every day	0.85	0.90	0.15	0.72
Voted 2010	0.77	0.66	−0.24	−1.12
Romney voter	0.24	0.22	−0.07	−0.34
Hispanic threat	0.06	0.05	−0.07	−0.33
Age	44.66	40.43	−0.35	−1.63
Residency year	8.22	7.07	−0.19	−0.91
College	0.89	0.86	−0.06	−0.30
Male	0.60	0.60	−0.01	−0.03
Hispanic	0.03	0.05	0.11	0.58
White	0.91	0.83	−0.24	−1.25
Income	146,236	140,103	−0.08	−0.42
n	117	103		

*Difference in standardized units.
[†]Mean response values for pretreatment variables accounting for stratification into train stations. All variables are 0 and 1 variables, except for Hispanic threat, which is a seven-point scale indicating how threatening respondents find Hispanics, recoded 0–1; residency, which is measured in years; and income, which is annual income in dollars.

The locations used were the Franklin Line, running southwest from Boston toward Franklin, MA, and the Framingham/Worcester Line, running west from Boston toward Worcester, MA. So that confederates could visit several stations in the same day, I chose stations that were clustered within lines, thus allowing the confederates to ride the train inbound and get on and off the train to treat subsequent stations. First, stations were eliminated that were either outside of Massachusetts or within the MBTA "Zone 1" and "Zone 1A," which is roughly the city of Boston and the most immediate suburbs. Then stations were eliminated that were within a Census Designated Place or Census Tract that was over the 66th percentile Hispanic of the eligible stations (this was over 6%). This aspect means that every place selected was below the US Census Designated Place mean percent Hispanic of 10%.

Matched pairs of trains always had at least one train separating them in time for the purpose of attempting to minimize contamination between treatment and control units; this was in case, for example, a passenger normally at a treatment unit missed his or her usual train and instead caught the next train arriving at the station. This sort of contamination would lead to a diminishing of difference between treatment and control. In analysis, I can attempt to eliminate this contamination by removing subjects who said they had missed their regular train during the period of treatment. Doing so makes no substantive difference in the results.

Each confederate was provided with a GPS device so I could be sure they were present at the stations at the designated times. The confederates were also asked to report on his or her experiences after riding the train each day. Monitors were also occasionally sent to check that the confederates were performing their roles as instructed. These monitors remained anonymous to the confederates. The confederates were hired in pairs, so that they would be familiar with each other and, therefore, be more likely to be comfortable with each other and to speak with each other. They were instructed to arrive at the train station at a time well before the arrival of the targeted train, wait on the platform, and get on the next train when it arrived.

The confederates were six men and one woman. All were Mexican nationals living in the United States on visas. Only one had lived in the United States longer than 1 y. The confederates were all between 21 and 23 y old, except for one who was 29 y. All had at least some college education, with one being a 4-y college graduate and another having a postgraduate degree. One confederate was a member of two different pairs.

Survey Recruitment. Five days before beginning treatment, members of my research team visited each treatment unit and distributed invitations to persons waiting for the train. The invitation was to take a survey about their political opinions. The invitation consisted of a document with a URL and two Visa gift cards. The document stated that the first card had $5 on it that they were free to use and that the second card had $5 that would be activated once they visited the URL and completed the survey. (A pilot invitation revealed that subjects were just as likely, if not more likely, to respond to a $5 inducement as to a $10 or $15 inducement.)

Treatment Group Balance. The survey collected background characteristics on each subject (Table 2). An omnibus test of balance correcting for the clustered assignment (46) demonstrates that randomization was successful ($c^2 = 13.4$ on 14 df, $P = 0.492$). A slightly greater percentage of white persons are in the control group than in the treatment group. Imbalance on a single characteristic is to be expected by chance alone, but because the experiment deals with attitudes associated with ethnicity as an outcome variable, imbalance on this variable is notable. Controlling for this variable (39) makes no substantial difference in the results.

Estimation of Experimental Effects. To estimate uncertainty in the quantities reported in Table 1, I use the method of Randomization Inference to test the probability of obtaining the observed ATE given all of the possible permutations of randomization, accounting for the matched pair design and clustering of respondents in stations and controlling for the MBTA line on which the station is located. The reported *P* values are generated by one-tailed tests. Two-tailed tests generated *P* values of 0.016, 0.031, and 0.540.

Reported results include responses in only 16 stations within eight matched pairs, rather than the original 18 stations, because one pair had such low distribution and response rates that no subjects from the treatment group entered T2. As such, the handful of subjects in the control condition did not have a matched pair, so both stations were dropped. The matched-pair design ensures that dropping these subjects does not bias estimation (47). I also estimated treatment effects without dropping this pair and the substantive results remain unchanged.

There were 20 subjects who entered the survey after the beginning of the treatment and were therefore discarded. These subjects can be used for a robustness check: They were exposed to the treatment before completing T1, so their responses in T1 should be more exclusionary than those of other subjects. This prediction is supported by the data, although with only 20 subjects, the differences of means contain considerable uncertainty and, of course, these subjects were not randomly assigned, so inferences should be made with caution.

Table 3. Completion percent by treatment condition

Control or treatment	*n* distributed	T1 (%)	T2* (%)
Control	100[†]	48.3	58.1
	(242)[‡]	(117)	(68)
Treatment	100	42.7	53.4
	(241)	(103)	(55)

*T2 percents represent the percent of T1 completes that also completed T2.
[†]Cell entries are the percent completed at each survey wave.
[‡]*n* in each cell are in parentheses.

I can check for realized heterogeneity by confederate pair in the treatment effects by subsetting by confederate pair or by interacting treatment and confederate pair in a regression model. The confederate pairs had very similar treatment effects.

Confidence intervals for the time effects reported in Fig. 2 are constructed by drawing the 2.5% and 97.5% quantiles from the randomization distribution (39).

Survey Attrition. A similar number of invitations were distributed at treatment and control units, a similar percentage of invited persons in treatment and control units completed T1, and a similar percentage of T1 completers in each condition completed T2 (Table 3). There is a slightly higher completion rate in T2 in the control group. Of respondents who completed T1, about four fewer individuals completed the T2 survey in the treatment condition than would have completed it if the two conditions had equal attrition. This imbalance can happen by chance in a small sample and the relationship between treatment assignment and missingness is not statistically significant. Of course, this test of significance may be statistically underpowered and it is important to ensure that attrition is not driving the experimental results. Fortunately, because the design included a pretest element, we can check to see if subjects fail to complete the second survey in a way that is "as if random." I performed an omnibus test (46) comparing subjects who completed T2 to those who did not. The propensity is unrelated to most background characteristics, including measured political attitudes and responses to the dependent variables of interest. Attrition is significantly related to three characteristics, all of which might be expected: People who failed to complete the second round were less likely to be regular MBTA riders, less likely to have voted in 2010, and younger. It is not immediately obvious why any of these variables would be correlated with potential outcomes; however, I cannot be sure, so I used simulations to test the robustness of the results in light of the slight imbalance in attrition across assignment.

I simulated additional datasets consisting of the observed data and four additional respondents in the treatment group: first with all four additional respondents having maximum positive changes in their responses to policy variables between T1 and T2 (responding 1 on all questions in T2) and second by all four respondents having maximum negative changes between T1 and T2 (responding 0 on all questions in T2). Using these extreme values, I can be sure that the hypothetical ATE that I would have obtained if these four individuals had completed both surveys would be somewhere in the range of the simulated values (48). I simulated draws of individuals from different stations to account for the clustering of respondents into stations. I performed Randomization Inference on each of 1,000 simulated datasets that include the hypothetical individuals with extreme values. For the first two questions in Table 1, the averaged ATEs show that even in the extreme negative case, where all attrited individuals' responses shifted all of the way in the nonexclusionary direction, the expected ATE is still a shift in the exclusionary direction, although obviously with smaller effect sizes [ATE (SE) for questions in Table 1: 0.052 (0.003), 0.038 (0.003), -0.018 (0.005)].

ACKNOWLEDGMENTS. I thank Adam Glynn, Anthony Fowler, Victoria Shineman, and Kris-Stella Trump for helpful comments on research design; Andrew Blinkinsop, Chelsea Celistan, Alexander Sahn, and especially Dana Higgins and Ola Topczewska for research assistance; Xanni Brown, Cody Dean, and Dasha Slavina for logistical assistance; Helene Valencia for translation services; Lawrence D. Bobo and participants at the 2012 Racial Attitudes and Identities Network meeting at Harvard University; and the confederates: Fabian Gil Bedoya, Carolina Cubedo, Jose Leonardo Cubedo, Victor Figueroa, Andres Garcia, Hugo Palafox-Carlos, and Rafael Salomon. This study was supported in part by Bruce Western and the Malcolm Weiner Center for Social Policy at Harvard University and the Behavioral Laboratory in Social Sciences at Harvard University.

1. Forbes HD (1997) *Ethnic Conflict: Commerce, Culture, and the Contact Hypothesis* (Yale Univ Press, New Haven).
2. Putnam RD (2007) E Pluribus Unum: Diversity and community in the twenty-first century: The 2006 Johan Skytte Prize Lecture. *Scand Polit Stud* 30(137):137–174.
3. Habyarimana J, Humphreys M, Posner DN, Weinstein JM (2007) Why does ethnic diversity undermine public goods provision? *Am Polit Sci Rev* 101(04):709–725.
4. Couzin ID, et al. (2011) Uninformed individuals promote democratic consensus in animal groups. *Science* 334(6062):1578–1580.
5. Lim M, Metzler R, Bar-Yam Y (2007) Global pattern formation and ethnic/cultural violence. *Science* 317(5844):1540–1544.
6. Vanhanen T (1999) Domestic ethnic conflict and ethnic nepotism: A comparative analysis. *J Peace Res* 36(1):55–73.
7. Dahl RA (1961) *Who Governs?: Democracy and Power in an American City* (Yale Univ Press, New Haven).
8. Alesina A, Spolaore E (2003) *The Size of Nations* (MIT Press, Cambridge, MA).
9. Pettigrew TF, Tropp LR (2006) A meta-analytic test of intergroup contact theory. *J Pers Soc Psychol* 90(5):751–783.
10. Sober E, Wilson DS (1998) *Unto Others: The Evolution and Psychology of Unselfish Behavior* (Harvard Univ Press, Cambridge, MA).
11. Bowles S (2009) Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? *Science* 324(5932):1293–1298.
12. Tajfel H (1969) Cognitive aspects of prejudice. *J Biosoc Sci* 25(4, Suppl 1):173–191.
13. MacLean RC, Gudelj I (2006) Resource competition and social conflict in experimental populations of yeast. *Nature* 441(7092):498–501.
14. Facchini G, Mayda AM (2009) Does the welfare state affect individual attitudes toward immigrants? Evidence across countries. *Rev Econ Stat* 91(2):295–314.
15. Posner DN (2004) The political salience of cultural difference: Why Chewas and Tumbukas are allies in Zambia and adversaries in Malawi. *Am Polit Sci Rev* 98(4): 529–545.
16. Oliver JE (2010) *The Paradoxes of Integration: Race, Neighborhood, and Civic Life in Multiethnic America* (Univ of Chicago Press, Chicago).
17. Falk A, Heckman JJ (2009) Lab experiments are a major source of knowledge in the social sciences. *Science* 326(5952):535–538.
18. Gerber AS, Green DP, Kaplan EH (2004) in *Problems and Methods in the Study of Politics*, eds Shapiro I, Smith R, Masoud T (Cambridge Univ Press, London), pp 251–273.
19. Sampson RJ (2008) Moving to inequality: Neighborhood effects and experiments meet social structure. *Am J Sociol* 114(1):189–231.
20. Sears DO (1983) The person-positivity bias. *J Pers Soc Psychol* 44(2):233–250.
21. Allport GW (1954) *The Nature of Prejudice* (Addison-Wesley, Cambridge, MA).
22. Aberbach J, Walker JL (1973) *Race in the City* (Little, Brown, Boston).
23. Edsall TB, Edsall MD (1992) *Chain Reaction: The Impact of Race, Rights, and Taxes on American Politics* (Norton, New York).
24. Rieder J (1985) *Canarsie: The Jews and Italians of Brooklyn Against Liberalism* (Harvard Univ Press, Cambridge).
25. Hopkins DJ (2010) Politicized places: Explaining where and when immigrants provoke local opposition. *Am Polit Sci Rev* 104(40):40–60.
26. Key VO (1949) *Southern Politics in State and Nation* (Knopf, New York).
27. Sidanius J, Levin S, Van Laar C, Sears D (2008) *The Diversity Challenge* (Russell Sage, New York).
28. Duckitt J (2003) in *Oxford Handbook of Political Psychology*, eds Sears DO, Huddy L, Jervis R (Oxford Univ Press, New York), pp 559–600.
29. Hoffman E, McCabe K, Smith VL (1996) Social distance and other-regarding behavior in dictator games. *Am Econ Rev* 86(3):653–660.
30. Tajfel H, Billig MG, Bundy RP, Flament C (1971) Social categorization and intergroup behaviour. *Eur J Soc Psychol* 1(2):149–178.
31. Green DP, Wong JS (2009) in *The Political Psychology of Democratic Citizenship*, ed Borgida E (Oxford Univ Press, New York), pp 559–600.
32. Brewer M, Miller N (1984) *Groups in Contact: The Psychology of Desegregation*, eds Brewer M, Miller N (Academic, San Diego), pp 281–302.
33. Pettigrew TF (1997) Generalized intergroup contact effects on prejudice. *Pers Soc Psychol Bull* 23(2):173–185.
34. Pettigrew TF (1998) Intergroup contact theory. *Annu Rev Psychol* 49(1):65–85.
35. Sacerdote B (2001) Peer effects with random assignment: Results for Dartmouth roommates. *Q J Econ* 116(2):681–704.
36. Boisjoly J, Duncan GJ, Kremer M, Levy DM, Eccles J (2006) Empathy or antipathy? The impact of diversity. *Am Econ Rev* 96(5):1890–1905.
37. Shook NJ, Fazio RH (2008) Interracial roommate relationships: An experimental field test of the contact hypothesis. *Psychol Sci* 19(7):717–723.
38. Sears DO (1986) College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *J Pers Soc Psychol* 51(3):515–530.
39. Gerber AS, Green DP (2012) *Field Experiments: Design, Analysis, and Interpretation* (Norton, New York).
40. Welch S, Sigelman L, Bledsoe T, Combs M (2001) *Race and Place: Race Relations in an American City* (Cambridge Univ Press, New York).
41. Knowles ED, Peng K (2005) White selves: Conceptualizing and measuring a dominant-group identity. *J Pers Soc Psychol* 89(2):223–241.
42. Bobo L (1983) Whites' opposition to busing: Symbolic racism or realistic group conflict? *J Pers Soc Psychol* 45(6):1196–1210.
43. McGarty C, Penny REC (1988) Categorization, accentuation and social judgement. *Br J Soc Psychol* 27(2):147–157.
44. McGuire WJ, McGuire CV, Child P, Fujioka T (1978) Salience of ethnicity in the spontaneous self-concept as a function of one's ethnic distinctiveness in the social environment. *J Pers Soc Psychol* 36(5):511–520.
45. US Census Bureau (2012) *Population Projections*, Available at www.census.gov/population/projections. Accessed June 8, 2012.
46. Hansen BB, Bowers J (2008) Covariate balance in simple, stratified and clustered comparative studies. *Stat Sci* 23(2):219–236.
47. Imai K, King G, Nall C (2009) The essential role of pair matching in cluster-randomized experiments, with application to the Mexican Universal Health Insurance Evaluation. *Stat Sci* 24(1):29–53.
48. Manski CF (1990) Nonparametric bounds on treatment effects. *Am Econ Rev* 80(2): 319–323.