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;

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 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.

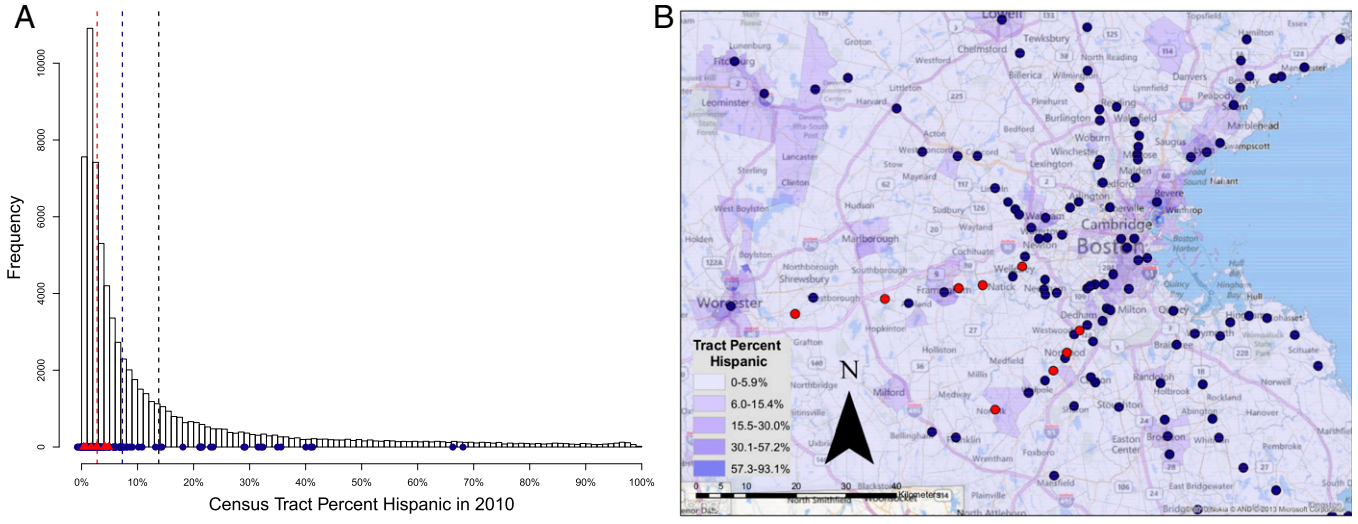

Fig. 1. Distribution of Hispanic persons (of any race) by Census Tract in the United States and in the experiment location in 2010. (*A*) Frequency of Census Tracts by percent Hispanic in the United States. Purple dots are percent Hispanic in Census Tracts containing an MBTA station and red triangles 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 bein[redacted] like them and bein[redacted] Latino."

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).