Supporting Information for "Changing climates of conflict"

1 Materials and Methods

Roots intervention description

Below we summarize the Roots intervention curriculum, the anti-conflict curriculum developed for the purposes of the study, by Allison Bland, Jennifer Dannals, Ariel Domlyn, and Laura Spence-Ash in collaboration with the first two authors. This summary covers the intervention curriculum's guiding principles, with references to example activities. The full intervention curriculum, with a complete activity explanation and a script for each meeting with students, can be downloaded on the website: Roots intervention curriculum link.

Each principle described below guides activity and discussion topic choice throughout the curriculum. Points 1 - 6 follow a rough chronological introduction of the program's principles; for example, students spend more time identifying topics and examples of conflict in their schools in the beginning of the program, and more time reaching out to other students toward the middle and end of the curriculum.

Logistics: students meet every other week with an adult leader in a mixed-gender, mixed-grade group to discuss and participate in activities. The curriculum contains ten meetings that took place in the study over the course of a school year. Each meeting lasted the course of a school period (typically, 45 to 60 minutes).

Curriculum guiding principles:

1. Students identify the issues: Roots meetings are safe spaces where students can speak openly and honestly without fear of reporting to other students or adults. Roots students identify areas for improvement that are important and noticeable to them about how students treat each other in the school. Students think about particular scenarios that happen in their school to learn how to observe things they

see going on and connect those observations to opportunities for change. Students are reminded that they are the experts in their school.

- Example Activities: Make Change, Xtranormal

2. Students generate possible solutions: Roots students generate suggestions for addressing areas for improvement in how students treat each other, starting with how to treat friends. They are guided to think through specifics of something they might do when they see certain things, and what they might encourage their friends to do. Students are encouraged to select behaviors that are comfortable and natural for them. Students are encouraged to feel influential and able to affect what happens at their school.

    - Example Activities: Bank of Behaviors, Flowchart, Role-play

3. Provide opportunities for student action: Activities are set up to provide students with specific actions to take in response to something they see happening in their school among students, or to start a change. They are in charge of creating the materials that are part of their activities. Students feel ownership of their school and the climate at their school as they take on the Roots identity and form relationships with other students in the program.

    - Example Activities: Weekly challenges, Pay it Forward, Hashtag posters, Orange wristbands

4. Make student initiatives visible to others: It is very important to demonstrate to the rest of the school the activities of the Roots students so that the Roots students and their actions are visible to others. Enable students to describe what the group is about, and what they themselves are doing to make change. Help students practice explaining what the group is about to others so other students understand.

    - Example Activities: Roots Day, morning announcements, posters at the school, Make Change box, Taboo game

5. Students use online platforms and techniques to collaborate and reach others: Much of student initiative, creativity, and connection occur online. This program can use how students actually interact to help students reflect on what happens at their school, and to use social media to publicize their activities and viewpoints. Students are interested in learning about how to use new online platforms to create messages. They like to be able to make something that “looks good,” or “professional,” in a way that might not be as easy offline. But they want to know who they are spreading messages to – this helps them feel in control of connecting on platforms they already use and have certain identities and reputations on.
    - Example Activities: Roots website, students take pictures of their work to post to their Instagram accounts or Facebook, hashtag posters (connection to online activities), making videos to post online

Average attendance rates for each meeting for each of the first nine intervention meetings (only seven schools held ten total meetings):

Meeting Number	Attendance Rate
1	75.3%
2	64.6%
3	62.6%
4	55.2%
5	54.7%
6	55.3%
7	60.0%
8	57.4%
9	70.1%

Study design

In this section, we summarize material seen elsewhere in the paper and in the Supplementary Information section.

The research design for this study was reviewed and approved by Princeton University's Institutional Review Board (protocol #4941). In collaboration with the New Jersey Department of Education, the researchers sent an email to all public middle schools in New Jersey inviting them to apply to participate in a cost-free research intervention. Over 110 middle schools responded. One factor leading to this enthusiastic response was a New Jersey anti-bullying law that took effect the year before the intervention, which mandated that schools provide anti-bullying programming. From this sample, 60 schools were selected on the basis of their geographic location and loose similarity to other schools. Prior to randomization, schools agreed to participate in all measurement aspects of the program, with a 50% chance of receiving the anti-conflict program (the treatment).

Schools were assigned to blocks of four, and randomized to receive the treatment or not within these blocks. Blocks were composed to maximize balance on the following variables: the latitude and longitude location of each school, the average school population as measured by the number of students who took the pre-randomization student surveys, the 5th, 6th, and 7th grade population during the year prior to the study (2011), the percentage of students identified as white, black, and hispanic, the percentage of students identified as having limited English proficiency, and the percentage of students receiving free or reduced lunch as identified by the New Jersey State Department of Education, and finally the average network clustering coefficient and network density calculated from student network data gathered in the pre-randomization student surveys. Four schools dropped out from the study before the point of randomization, which left the total sample at 56 schools and 24,191 students.

We now describe the algorithm that determined which students were eligible for randomization into the seed group. We first blocked on gender and grade within schools.

Student ids were ordered alphabetically within each gender-grade block. For each student with id i, we computed the indegree and closure associated with the $(i + k)$th student, where k is the number of students who failed to take the survey with id $< i$. Denote indegree as a_i and closure as b_i. The students were then blocked into those with the top half of b_i vs. the bottom half of b_i. The students in each of the gender-grade-b_i blocks with the largest values of a_i were considered eligible for randomization into the seed group, such that we had no fewer than 40 seed-eligible students (schools with fewer than 200 students), and no more than 64 seed-eligible students (schools with less than 550 students). For schools with between 200 and 550, we scaled linearly between 40 and 64 seed-eligible students, rounding to the nearest 4. Table S2 demonstrates the covariate profiles of eligible and ineligible students. All of our estimates are computed conditional on the composition of the seed-eligible groups. Among treated schools, half of the seed-eligibles were block-randomized into the seed group, using 4 blocks constructed using gender, grade, and b_i (using an R package, blockTools).

The research team designed the intervention (the full curriculum, which includes all source materials and proper attribution to contributors, is posted at http://njroots.princeton.edu/curriculum). To implement the intervention, a trained research assistant met with the school's student group (selected by the research team through a process described in the article text) every other week and led activities where students identified common conflict behaviors at their school. This bottom-up approach was designed to allow students to address conflicts specific to their school. The school's group of students were encouraged to become the public face of opposition to these types of conflict. For example, groups at each school compiled a list of conflict behaviors they could address, created hashtag slogans about those behaviors, and turned the slogans into online and physical posters. The students' photos were posted next to the slogan in order to create an association between the anti-conflict statement and each participating student's identity. In another activity, the school's student group gave an orange wristband with the intervention logo (a tree) as a reward to students who were observed engaging in friendly or conflict mitigating behaviors (over 2,500 wristbands were distributed and tracked). This intervention model can be likened to a grassroots campaign in which the seed students took the lead; notably, it lacked an educational or persuasive unit regarding adult-defined problems at their school.

Survey wording and school administrative record description

Descriptive norms scale (asked at pretest and posttest):

"How often do you see students…"
(0 = Never, 1 = 1-2 times / month, 2 = About 1 time /week, 3 = 2-3 times / week, 4 = Every day)

1. Messing with, trash talk, picking on
2. Gossip and rumors
3. Being mean to a girl because of what she does with boys
4. Saying someone is fat, ugly, making fun of clothes
5. Posting or texting something mean
6. Threatening, hitting, or pushing
7. Joking about race or ethnicity
8. Calling others gay
9. Reporting conflicts to adults
10. Staying out of conflict
11. Being friendly or nice with everyone
12. Speaking up for someone
13. Saying mean things about clothes (item added for wave 2 only)

Prescriptive norms scale

Prescriptive norms scale (asked at pretest and posttest, answered using pictorial scale featured above):
"How many students think it is. . . "

1. Good to be friendly and nice
2. Good to stay out of conflict
3. Not funny to mess with or pick on others
4. Not OK to say mean things to girl because of what she does with boys
5. Not funny to post online or text mean things
6. Not OK to use racial, ethnic jokes
7. OK to speak up for students
8. OK to tell adults about conflict
9. Not good to threaten, hit, punch
10. Not OK to gossip or spread rumors
11. Not funny to call someone gay
12. Not funny to say mean things about clothes (item added for wave 2 only)

Talking to friends item: "This year, friends talked with me about how to change student behavior" (Yes/No)
Orange wristband report item: "I wore an orange Roots wristband this past week" (Yes/No)

Codes for administrative disciplinary action (counts summed for each individual and school)

Reporting practices for disciplinary events differed across schools. Schools' reporting practices were determined before they were in contact with the research team and before treatment began, and was unrelated to treatment assignment. Codes below represent those used by schools; when schools coded incidents using more detailed categories, those codes were classified within these broader descriptive codes. Any missing data for a student or for an entire category of disciplinary action was treated in our analyses as zero. Seven of 56 schools did not provide any data, for varied reasons including physical document loss, failure to retain documents, and personnel turnover. Failure to provide data was unrelated to treatment status.

1. Inappropriate language towards student (written or verbal)
2. Verbal altercation with student
3. Bully/HIB
4. Making threat
5. Spreading rumor
6. Biased comment / slur (Race/gender/sexual orientation/etc.)
7. Inappropriate (nonverbal) gestures

8. General (e.g., being mean)
9. Physical aggression with student (violence, pushing, kicking, one way or two way, may or may not include the term fight or altercation)
10. Inappropriate physical contact (such as pantsing, spitting, etc.)
11. Horseplay (clear that it is physical joking, but still disruptive and thus punished)
12. Sexual harassment, inappropriate sexual conduct (inc. sexual behavior in school)
13. Inciting violence/planning fight
14. Inappropriate language towards student (written or verbal)
15. Verbal altercation with student
16. Bullying / H.I.B. report (H.I.B.: harassment, intimidation, and bullying New Jersey state report)
17. Making threat
18. Spreading rumor
19. Biased comment / slur (Race/gender/sexual orientation/etc.)
20. Inappropriate (nonverbal) gestures
21. General (e.g., being mean)
22. Physical aggression with student (violence, pushing, kicking, can be one way or two way, may or may not include the term fight or altercation)
23. Inappropriate physical contact (such as pantsing, spitting, etc.), NOT pushing/hitting/shoving/tackling
24. Horseplay (clear that it is physical joking, but still disruptive and thus punished)

25. Sexual harassment, inappropriate sexual conduct (inc. sexual behavior in school)

26. Inciting violence/planning fight

27. Number of H.I.B. incidents as offender

28. Number of H.I.B. incidents as target

Analyses

We now discuss our analysis of the experimental data. The randomization of schools and seeds facilitates the design-based evaluation of the causal effects of our experimental intervention. *Climate effects* are based on the between-school randomization to treatment and control, with linear regression of school-level outcomes on school-level assignment and covariates producing consistent estimates of average school-level causal effects. We fit the following linear regression model:

$$Y_{ij} = \alpha_j + \tau D_{ij} + \gamma C^E_{ij} + u_{ij} \tag{1}$$

where ij denotes the ith school in the jth block. Y_{ij} denotes school-level outcomes, α_j is a block-level fixed effect, D_{ij} is the school-level treatment assignment, C_{ij} is the composition of the seed group (percentage of seed group that is classified as a social referent), and C^E_{ij} is the composition of the seed eligible group (percentage of seed eligible group that is classified as a social referent). τ here is the average causal effect of treatment. We include C^E_{ij} as a covariate control here so as to maximize comparability with the interaction model in (4). In all cases, the model is fit with OLS regression with heteroskedasticity-robust standard errors to characterize uncertainty.

To characterize how the seed group's composition (i.e., percentage of referents) affects the efficacy of the program, we also use an interaction model.

$$Y_{ij} = \alpha_j + \tau_0 D_{ij} + \beta C_{ij} D_{ij} + \gamma C^E_{ij} + u_{ij}. \tag{2}$$

Here, assuming that the linear approximation for causal effect heterogeneity is valid, τ_0 is the average causal effect of treatment when the composition of the seed group is 0% referents and β represents the slope of the average causal effect with respect to group composition in percentage points (i.e., assuming that the slope is constant, the average causal effect of treatment at a given composition $c\%$ is $\tau_0 + \beta c$.) As $\mathrm{E}\left[C_{ij}\right] \propto C^E_{ij}$, controlling for C^E_{ij} ensures that any random deviations in C_{ij} from $\mathrm{E}\left[C_{ij}\right]$ may be interpreted causally. For both models, we also use pretreatment covariates (percentage white, percentage English-speaking, average house quality and, where available, pretreatment measurement of Y_{ij}) to improve efficiency by including them as linear controls in the regression model.

Social influence effects are based on the random assignment of treatment to seed eligibles and assessed by how seed students causally affect other students in their social network. We allow for each student i to have a vector of potential outcomes, which we denote by $(y_i(d_C), y_i(d_S), y_i(d_{NI}), y_i(d_I))$ associated with control exposure, school exposure, non-social referent exposure, and social referent exposure respectively. We seek estimates for the average potential outcomes $\mu(d_k) = \frac{1}{N}\sum_{i=1}^{N} y_i(d_k)$. An average unit-level causal effect is defined in terms of a difference between the average of students' potential outcomes under one exposure versus the average under another exposure, i.e., $\mu(d_k) - \mu(d_l)$, for some k, l.

For each student i, we have a vector of probabilities, characterized by the vector $(\pi_i(d_C), ..., \pi_i(d_I))'$. This vector tells us the probability of i being subject to each of the possible exposures in $\{d_C, ..., d_I\}$, $\pi_i(d_k) = \sum_{\mathbf{z}\in\Omega} \mathbf{I}(f(\mathbf{z}, \theta_i) = d_k) \Pr(\mathbf{Z} = \mathbf{z})$. Each component probability, $\pi_i(d_k)$, is, by definition, equal to the expected proportion of treatment assignments that induce exposure d_k for student i. We achieve an arbitrarily precise estimate of π_i via $J = 10,000$ simulated random assignments of $\mathbf{Z}$ according to the experimental design: $\hat{\pi}_i(d_k) = \mathbf{I}[\frac{1}{J}\sum_{\ell=1}^{J} f(\mathbf{z}^J, \theta_i) = d_k]$, where J is the number of simulated random assignments, and z^J denotes the Jth simulated random assignment vector. We then restrict ourselves to the subpopulation of students for which there exists some probability that students could receive all four exposures: $0 < \pi_i(d_k) < 1$ for $k \in \{C, S, NI, I\}$.

We use a covariate-adjusted inverse probability weighted regression estimator to estimate our average potential outcomes and therefore average causal effects. Randomization-based confidence intervals for average potential outcomes are obtained under a maintained hypothesis of constant effects. Specifically, we use 1,000 simulated randomizations (both randomization across schools and within seed-eligibles) to estimate the standard errors under the maintained hypothesis that, for all i, $y_i(d_k) = y_i(D_i) + \hat{\mu}_R(d_k) - \hat{\mu}_R(D_i)$. Wald-type confidence intervals and p-values are then constructed using a normal approximation.

Analysis of Climate Effects

Here, we briefly describe the estimation strategy for climate effects. We fit the following linear regression model:

$$Y_{ij} = \alpha_j + \tau D_{ij} + \gamma C^E_{ij} + u_{ij}, \tag{3}$$

where ij denotes the ith school in the jth block. Y_{ij} denotes school-level outcomes, α_j is a block-level fixed effect, D_{ij} is the school-level treatment assignment, C_{ij} is the composition of the seed group (percentage of seed group that is classified as a social referent), and C^E_{ij} is the composition of the seed eligible group (percentage of seed eligible group that is classified as a social referent).. τ here is the average causal effect of treatment. We include C^E_{ij} as a covariate control here so as to maximize comparability with the interaction model in (4). In all cases, the model is fit with OLS regression with heteroskedasticity-robust standard errors to characterize uncertainty.

To characterize how the seed group's composition (i.e., percentage of referents) affects the efficacy of the program, we also use an interaction model.

$$Y_{ij} = \alpha_j + \tau_0 D_{ij} + \beta C_{ij} D_{ij} + \gamma C^E_{ij} + u_{ij}. \tag{4}$$

Here, assuming that the linear approximation for causal effect heterogeneity is valid, τ_0 is the average causal effect of treatment when the composition of the seed group is 0% referents and β represents the slope of the average causal effect with respect to group composition in percentage points. (I.e., assuming that the slope is constant, the average causal effect of treatment at a given composition $c\%$ is $\tau_0 + \beta c$.) As $\mathrm{E}\left[C_{ij}\right] \propto C_{ij}^E$, controlling for C_{ij}^E ensures that any random deviations in C_{ij} from $\mathrm{E}\left[C_{ij}\right]$ may be interpreted causally.

We also use covariates to improve the efficiency of our estimate. This amounts to fitting the following model for main effects:

$$Y_{ij} = \alpha_j + \tau D_{ij} + \gamma C_{ij}^E + \boldsymbol{\theta}\mathbf{X}_{ij} + u_{ij}, \tag{5}$$

where $\mathbf{X}_{ij}$ is a vector of school-specific pretreatment covariates (percentage white, percentage English-speaking, average house quality and, where available, pretreatment measurement of Y_{ij}). We also use covariates to adjust our interaction model:

$$Y_{ij} = \alpha_j + \tau_0 D_{ij} + \beta C_{ij} D_{ij} + \gamma C_{ij}^E + \boldsymbol{\theta}\mathbf{X}_{ij} + u_{ij}. \tag{6}$$

Interpretation of the causal effects proceed as above; inclusion of covariates is solely for efficiency improvement.

Analysis of Social Influence Effects

Our formulations in this section draw heavily on results from Aronow [52], and some notation and language is adapted accordingly. We have a finite population of students indexed by $i = 1, ..., N$ on which our randomized experiment is performed. Define a treatment assignment vector, $\mathbf{z} = (z_1, ..., z_N)'$, where $z_i \in \{0, 1\}$ denotes whether or not student i is assigned to be a seed (i.e., is invited to participate in the anti-conflict group). An experimental design is a plan for randomly selecting a particular value of $\mathbf{z}$ with predetermined probability $p_{\mathbf{z}}$. Restricting our attention only to treatment assignments that can be generated by our experimental design, define $\Omega = \{\mathbf{z} : p_{\mathbf{z}} > 0\}$, so that $\mathbf{Z} = (Z_1, ..., Z_N)'$ is a random vector with support Ω and $\Pr(\mathbf{Z} = \mathbf{z}) = p_{\mathbf{z}}$. In this setting, treatment assignment is the random assignment of seeds both across and within schools according our hierarchical randomization (i.e., so that half of schools are assigned to control schools with no seeds).

An exposure mapping is a unit-specific function that maps an assignment vector, $\mathbf{z}$, and student-specific traits, θ_i, to the actual exposure that a student receives. Each of the distinct exposures are allowed to give rise to distinct potential outcomes for each student. In this setting, exposure will refer to the student's proximity to referent seeds and nonreferent seeds (we will formalize this shortly). Define $D_i = f(\mathbf{Z}, \theta_i)$ such that $\Pr(D_i = d) = \pi_i(d)$. We define the exposure mapping as a function, $f(\mathbf{z}, \theta_i)$, such that the parameter, θ_i, equals student i's row in the network adjacency matrix. Further let γ be an N-length vector denoting referent vs. nonreferent status, so that the jth element of γ takes on the value 1 if student j has an indegree in the top decile of student j's school, and 0 otherwise. The crossproduct of z_i and θ_i, $\mathbf{z}'\theta_i$, is equal to the number of student i's peers that are assigned to be seeds. Further $(\mathbf{z} \circ \gamma)'\theta_i$ is equal to the number of student i's referent peers that are assigned to be seeds. For convenience, denote s_i as an indicator for whether student i's school has been assigned to be treated.

The exposure mapping is:

$$f(\mathbf{z}, \theta_i) = \begin{cases} d_C \text{ (Control exposure) :} & (1 - s_i) = 1 \\ d_S \text{ (School exposure) :} & (1 - z_i)\mathbf{I}(\mathbf{z}'\theta_i = 0)s_i = 1 \\ d_{NI} \text{ (Non social referent seed exposure) :} & \mathbf{I}((\mathbf{z} \circ \gamma)'\theta_i = 0)s_i = 1 \\ d_I \text{ (Social referent seed exposure) :} & (1 - z_i)\mathbf{I}((\mathbf{z} \circ \gamma)'\theta_i > 0)s_i = 1 \\ d_D \text{ (Direct treatment condition) :} & z_i = 1 \end{cases}$$

Although we have enumerated five different types of exposure, we will not consider the directly treated seed condition exposure, d_D, and instead will consider only d_C, d_S, d_{NI} and d_I. Without loss of generality, the exposure mapping could be generalized to allow influence effects on seeds, and the results would be unchanged.

Estimation of Average Causal Effects

We allow for each student i to have a vector of potential outcomes, which we denote by $(y_i(d_C), ...y_i(d_D))$, that do not depend on the value of $\mathbf{Z}$. We seek estimates for $k \in \{C, S, NI, I\}$ (i.e., control exposure, school exposure, non-social referent peer exposure, social referent peer exposure) of the average potential outcome $\mu(d_k) = \frac{1}{N}\sum_{i=1}^{N} y_i(d_k)$. An average unit-level causal effect is defined in terms of a difference between the average of students' potential outcomes under one exposure versus the average under another exposure, i.e., $\mu(d_k) - \mu(d_l)$, for some k, l. In order to do so, we will use a variant of inverse probability weighting.

For each student i, we have a vector of probabilities, characterized by the vector $(\pi_i(d_C), ..., \pi_i(d_D))'$. This vector tells us the probability of i being subject to each of the possible exposures in $\{d_C, ..., d_D\}$, $\pi_i(d_k) = \sum_{\mathbf{z}\in\Omega} \mathbf{I}(f(\mathbf{z}, \theta_i) = d_k) \Pr(\mathbf{Z} = \mathbf{z})$. Each component probability, $\pi_i(d_k)$, is, by definition, equal to the expected proportion of treatment assignments that induce exposure d_k for student i. We can achieve an arbitrarily precise estimate of π_i via simulated random assignments of $\mathbf{Z}$ according to the experimental design: $\hat{\pi}_i(d_k) = \mathbf{I}[\frac{1}{J}\sum_{\ell=1}^{J} f(\mathbf{z}^J, \theta_i) = d_k]$, where J is the number of simulated random assignments, and z^J denotes the Jth simulated random assignment vector. Aronow [52] provides details on the rate of convergence of the estimated probabilities. In our study, we use $J = 10,000$ replicates of the administration of treatment assignments to estimate the $\pi_i(d_k)$ values.

We restrict ourselves to the subpopulation of students (n=2451) for which there exists some probability that students could receive all four exposures: $0 < \pi_i(d_k) < 1$ for $k \in \{C, S, NI, I\}$. When $\pi_i(d_k) = 0$ for some students, then the design-based – i.e., extrapolation-free – estimation of average potential outcomes and causal effects must be restricted to the subset of students for which $\pi_i(d_k) > 0$. We will elaborate on this point shortly.

We will use a covariate-adjusted inverse probability weighted regression estimator to estimate our average potential outcomes and therefore average causal effects. For covariate adjustment, define an estimated parameter vector associated with exposure condition d_k

$$\widehat{\xi}(d_k) = \arg\min_{\xi(d_k)} \sum_{i:D_i=d_k} \frac{1}{\pi_i(d_k)} \left[y_i(d_k) - g\left(\mathbf{x_i}, \xi(d_k)\right)\right]^2,$$

where $g(.)$ is the specification for the regression of $y_i(d_k)$ on $\mathbf{x_i}$. $\xi(d_k)$ is the parameter vector arising from an inverse probability weighted regression of outcomes on covariates among students in a given level of exposure. Then the regression estimator for the mean is $\widehat{\mu_R}(d_k) =$

$$\frac{1}{N_s} \sum_{i:\pi_i(d_k)>0, \forall k \in \{C,S,NI,I\}} \left[\mathbf{I}(D_i = d_k) \frac{y_i(d_k) - g\left(\mathbf{x_i}, \widehat{\xi}(d_k)\right)}{\hat{\pi}_i(d_k)} + g\left(\mathbf{x_i}, \widehat{\xi}(d_k)\right)\right],$$

where N_s denotes the number of students such that $0 < \pi_i(d_k) < 1$ for all i and for $k \in \{C, S, NI, I\}$. Note that if any students were to have $\pi_i(d_k) = 0$ in the population under study, then the estimator would be undefined by virtue of division by zero. This implies that the subpopulation under study is restricted to those students who are connected to at least one seed-eligible non-social referent and at least one seed-eligible social referent. Thus our estimates of social influence effects are limited to this subpopulation, which constitutes 2,451 students.

Consistency and asymptotic normality directly follow from the law of large numbers and a Lindeberg-type central limit theorem given a growing sequence of independent schools of finite size, as holds by construction given our between-school randomization. Estimation of average causal effects follows using a plug-in estimate, $\hat{\mu}_R(d_k) - \hat{\mu}_R(d_l)$. Randomization-based confidence intervals for average potential outcomes are obtained under a maintained hypothesis of constant effects. Specifically, we use 1,000 simulated randomizations (both randomization across schools and within seed-eligibles) to estimate the standard errors under the maintained hypothesis that, for all i,

$y_i(d_k) = y_i(D_i) + \hat{\mu}_R(d_k) - \hat{\mu}_R(D_i)$. Aronow [52] derives the consistency of a variance estimator under this maintained hypothesis. Wald-type confidence intervals and p-values are then constructed using a normal approximation.

Network stability

Our inter-wave stability is typical of longitudinal social network data, given that our measurements were nearly nine months apart. 42.2% of reported ties in the (pretest) Wave 1 were persistent when we remeasured the network in the (posttest) Wave 2. Network stability is also often presented in the terms of a Jaccard index. The Jaccard index is the number of stable ties divided by the sum of the numbers of stable ties, new ties, and lost ties. Our Jaccard index is 0.264.

For comparison, we note three studies conducted with (older) students as participants:

- Veenstra and Steglich (2011, p. 605) consider friendship networks measured in Glasgow secondary schools over three annual surveys, and have wave-over-wave Jaccard indices of 0.275 and 0.314, with the two-year index at 0.201.
- Schaefer, Haas and Bishop (2012) consider high school students in the AddHealth data and find that friendship networks measured one year apart have a Jaccard index of 0.24.
- Weng, Hachen and Lizardo (2013) consider friendship networks of undergraduates measured one semester apart, and find wave-over-wave Jaccard indices of 0.41, 0.35, 0.41.

We also investigated the properties of the network measurement of 5th graders. Restricting our attention to the 1,968 5th graders in our experimental population (8.1% of the total study population), 41.9% of ties in Wave 1 were persistent in Wave 2, with a Jaccard index of 0.257. This result suggests no systematic difference between 5th graders and the rest of the population.

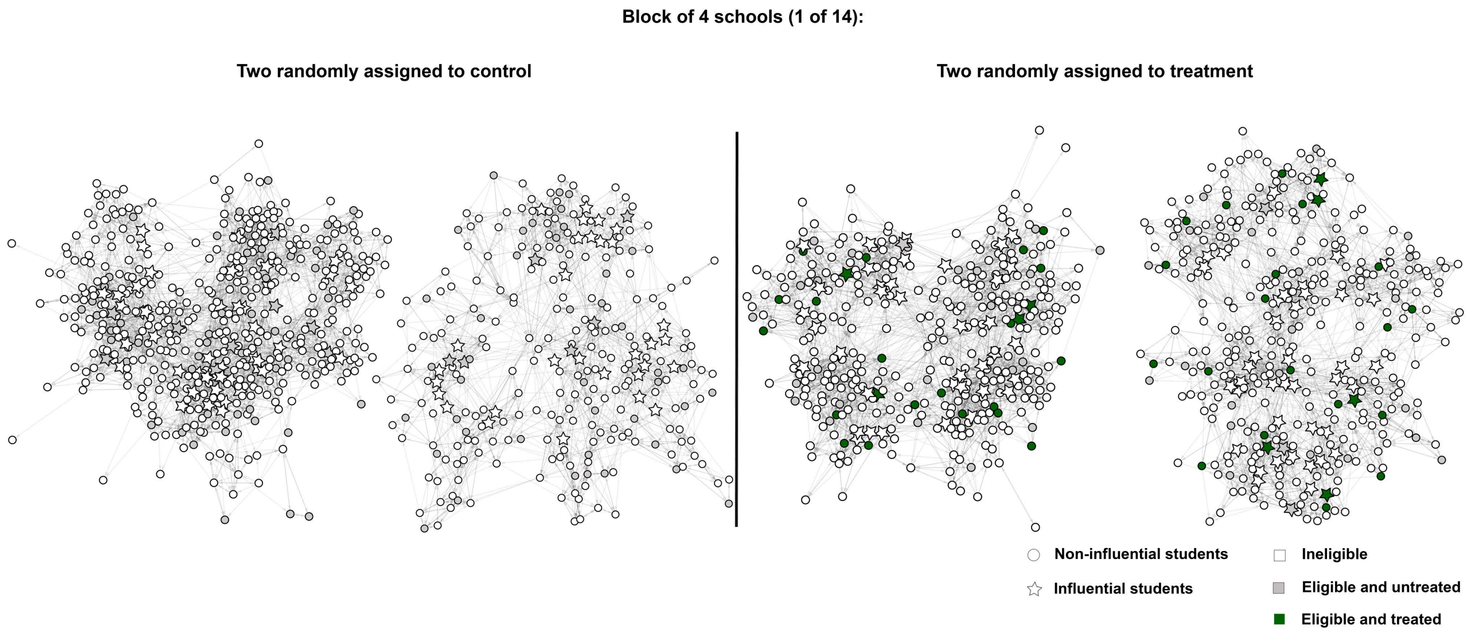
Descriptive statistics

Tables S1 and S2 provide descriptive statistics for our sample of schools and of students, respectively. We provide information on students' socioeconomic status, social status, activities and demographics because they represent the standard array of comparisons in much social influence and educational research. Many of them are commonly used in research with schools because the information is unobtrusively observed by schools (e.g., schools collect data on gender, parental occupation, and socioeconomic indicators like school lunch). Thus, for policy reasons, these covariates represent a way for researchers and practitioners involved in schools to understand the characteristics of social referents along these common and easily accessed dimensions. Some of our questions measure these constructs, such as socioeconomic status, using adolescent-friendly language that we piloted prior to our study with sample students drawn from the same geographic area and age group as our study participants. We used these new question wordings (e.g., in the case of socioeconomic status, we asked whether students' friends compliment the house where they live: "people say I have a nice house") as an assessment of these standard constructs.

2 Figures and legends

Experimental design

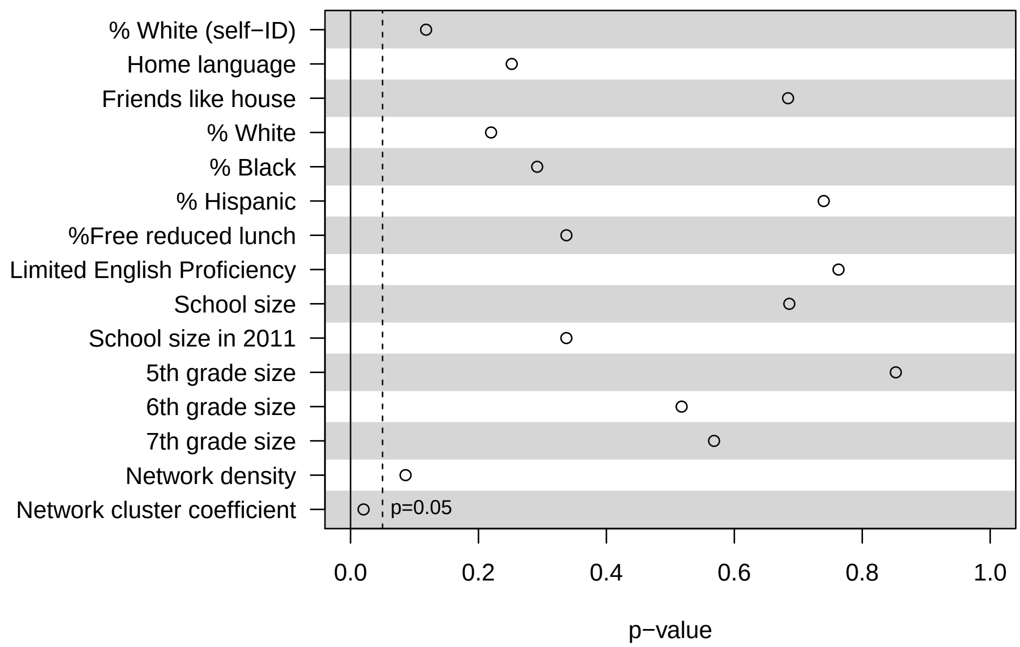
Fig. S1. Four school networks illustrate the experimental design.



The 56 schools were first blocked into groups of four by pretreatment school level variables including size and student demographics (see SI). Within each of the 14 school blocks, two schools were assigned to control (the two school networks on the left), and two to treatment, the anti-conflict intervention (the two school networks on the right). The nodes depict the population of students at the school. Grey nodes are the "seed eligible" students, selected by a deterministic algorithm and comprising a representative sample of the school. Seed eligibles were blocked into four equally sized groups based on pretreatment characteristics including gender and grade, and half of the students in each block were randomly assigned to be treated (part of the anti-conflict program) if they were in a treatment school; dark green nodes are treated seed eligibles. Nodes in the shape of stars are the social referents, who are in the top 10% of their school in connections reported by other students in the pre-randomization survey.

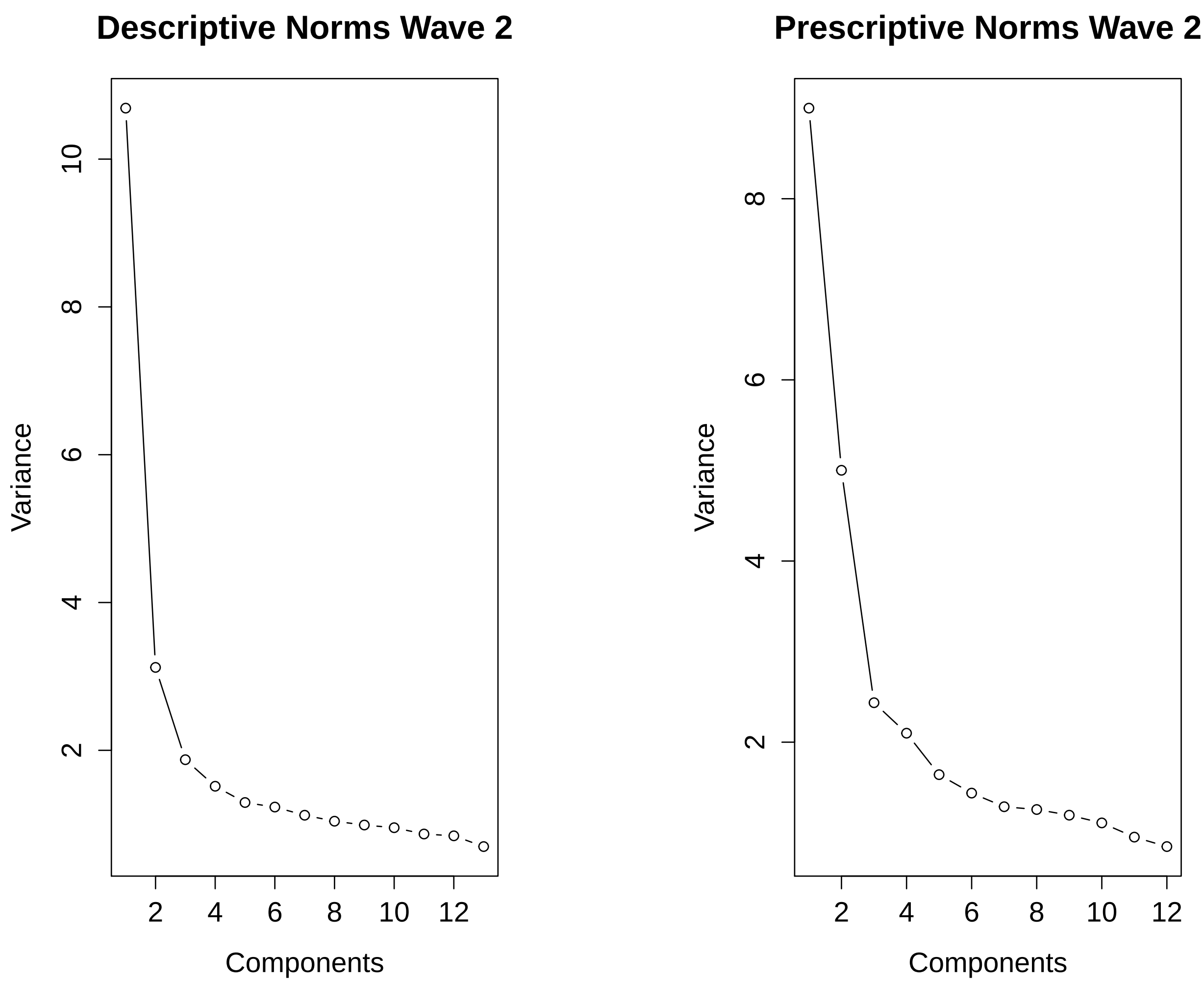
Descriptive statistics for schools

Fig. S2.Treatment-control balance across pretreatment covariates.



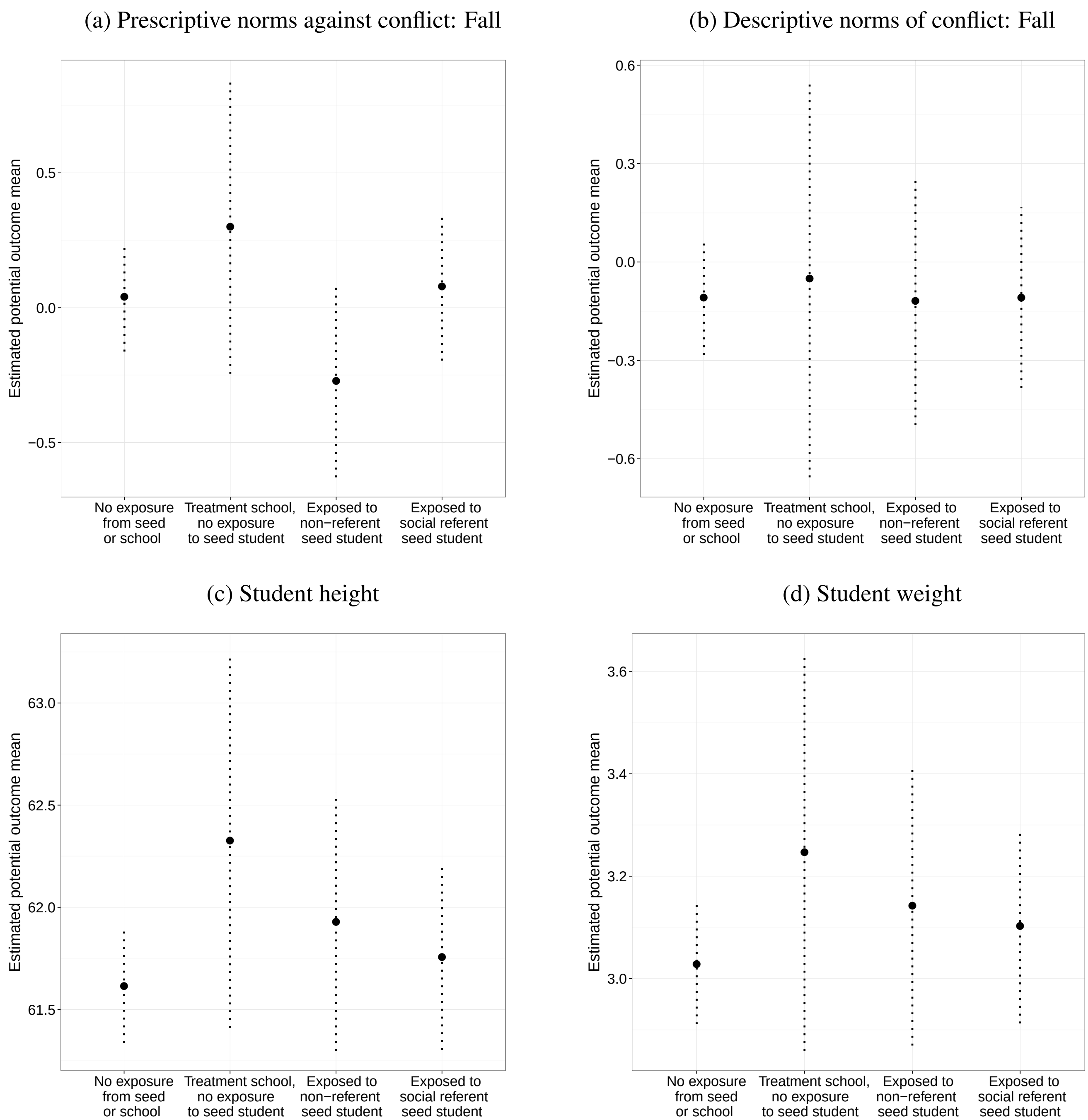
Each row of the figure represents the p-value against the null hypothesis that the given pretreatment school-level covariate is balanced across treatment and control schools. To compute these p-values, we use the school-level treatment assignment to predict the 15 pretreatment school-level covariates as outcomes in our regression specification. The regression follows the same specification used in the main text for our effect estimates, specifically, including fixed effects for school block, a control for average group composition, and robust standard errors. Under randomization, we should expect to see no systematic relationship between the school-level treatment assignments and these pre-treatment covariates–i.e., we should see no evidence of systematic imbalances. Indeed, we find no evidence of any systematic relationship between school treatment assignment and pretreatment covariates. Only one of the 15 covariates (network cluster coefficient) has a significant imbalance at the 0.05 level, and does not survive a Bonferroni correction. See Table 1 for variable descriptions.

Fig. S3. Scree plots for norm index from student surveys



Robustness checks: Placebo tests and alternative specifications

Fig. S4. No social influence effects for pretreatment or student individual characteristics

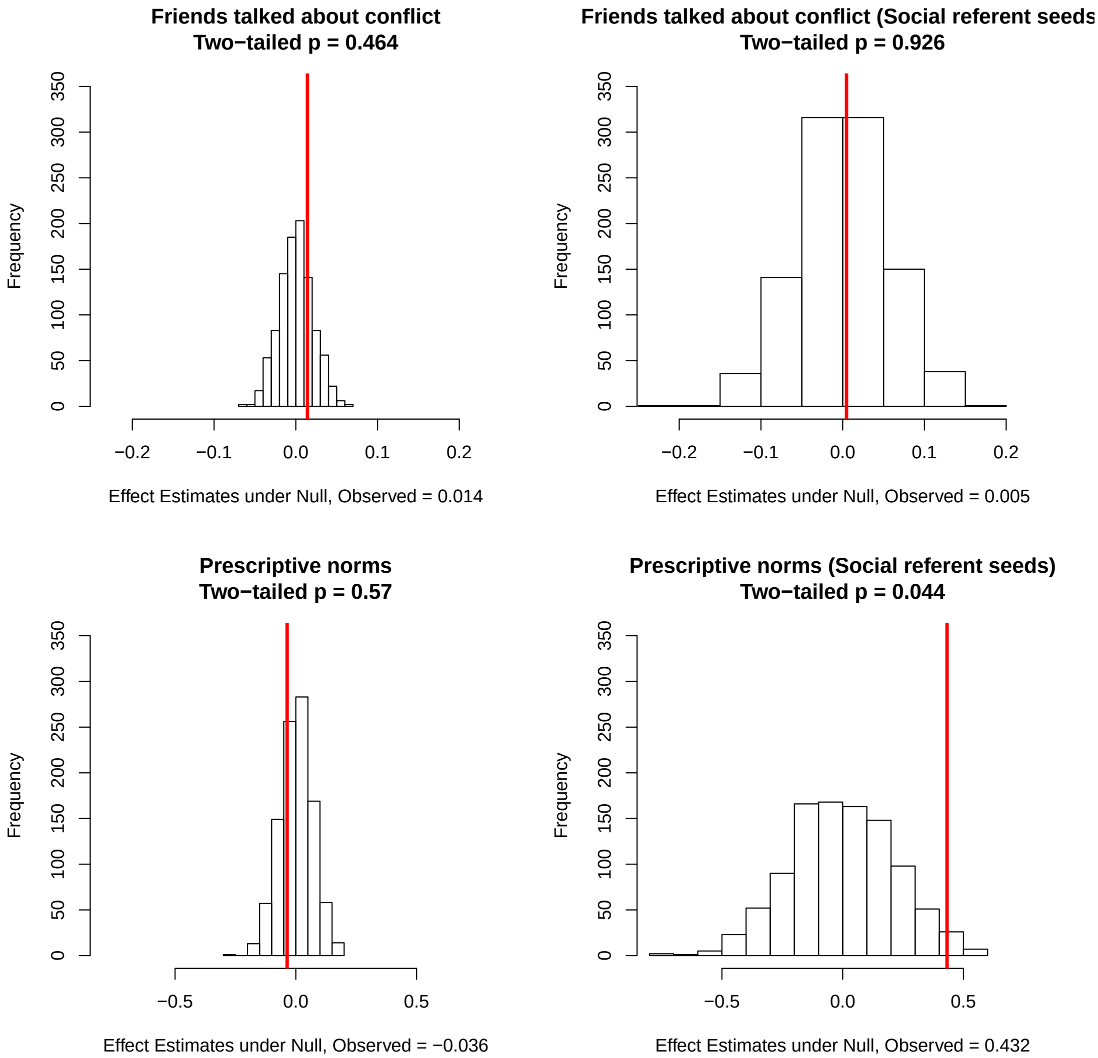


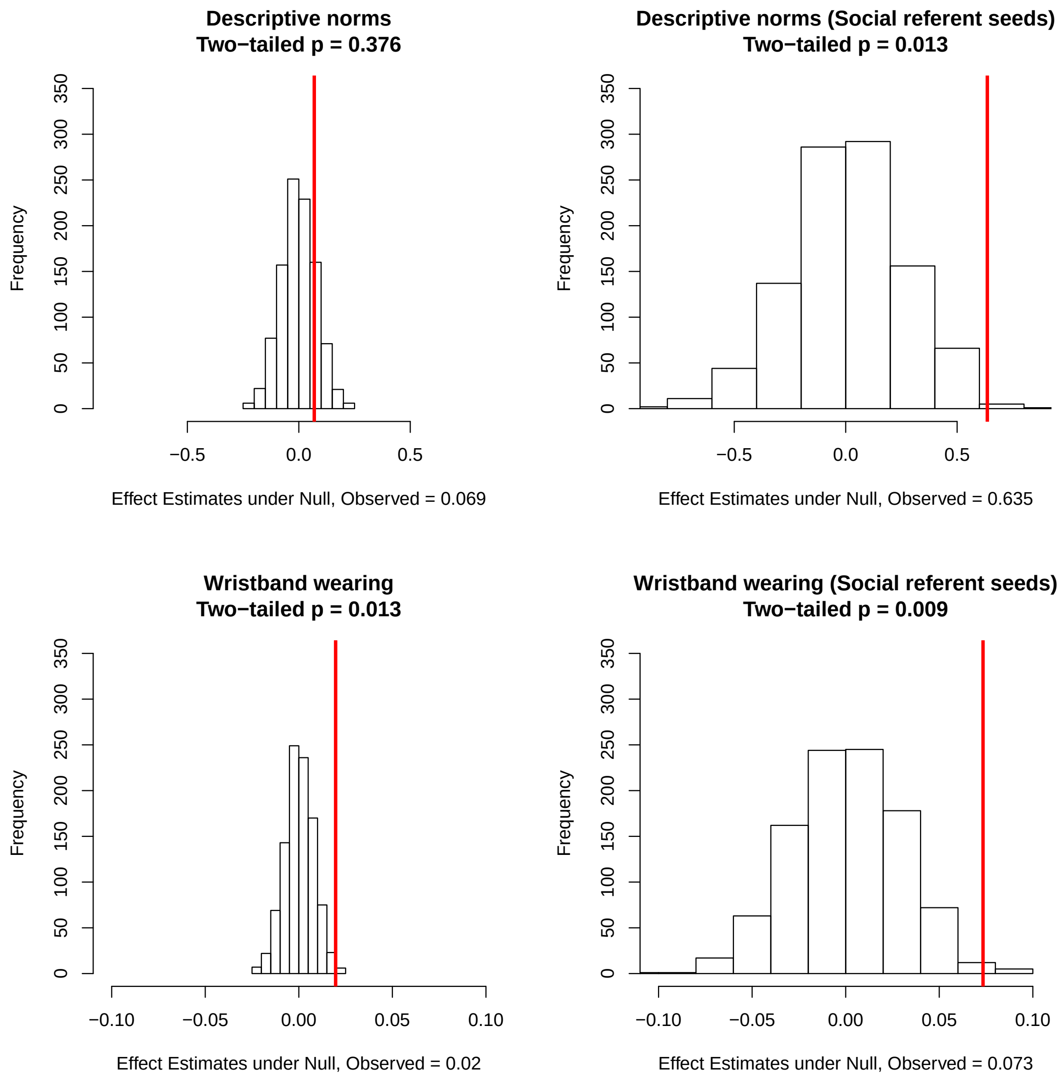
As an additional robustness check of our estimated network effects, we employed a variant of the method used by (*15*) to assess within-school social influence effects. We restrict our attention to treated schools, comparing the mean of outcomes of peers of seeds to the mean of outcomes of peers of seed-eligibles who were not selected. This difference-in-means serves as our test statistic.

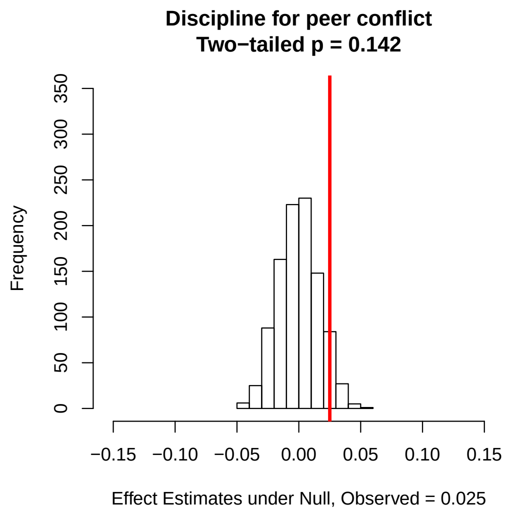
To compute a p-value against the sharp null hypothesis of no effect, we use randomization inference with 1,000 simulated permutations of treatment assignment, holding the network fixed. The proportion of randomizations such that the absolute value of the simulated difference-in-means is greater than or equal to the absolute value of the observed difference-in-means serves as our estimate of a two-tailed p-value. These results are visualized in the first column of the five graphs presented, where the histograms show the frequency of simulated test statistics, and the vertical lines show the observed test statistics.

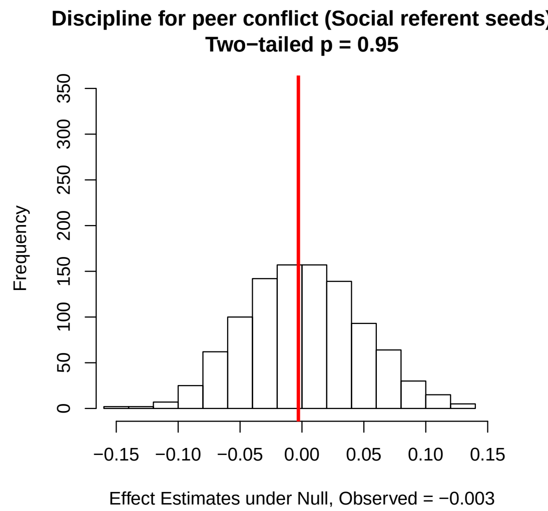
We also performed the same analysis by the mean of outcomes of peers of social referent seeds to the mean of outcomes of peers of social referent seed-eligibles who were not selected. p-values are computed in the same manner. These results are visualized in the second column of the five graphs presented.

Fig. S5. Alternative specification



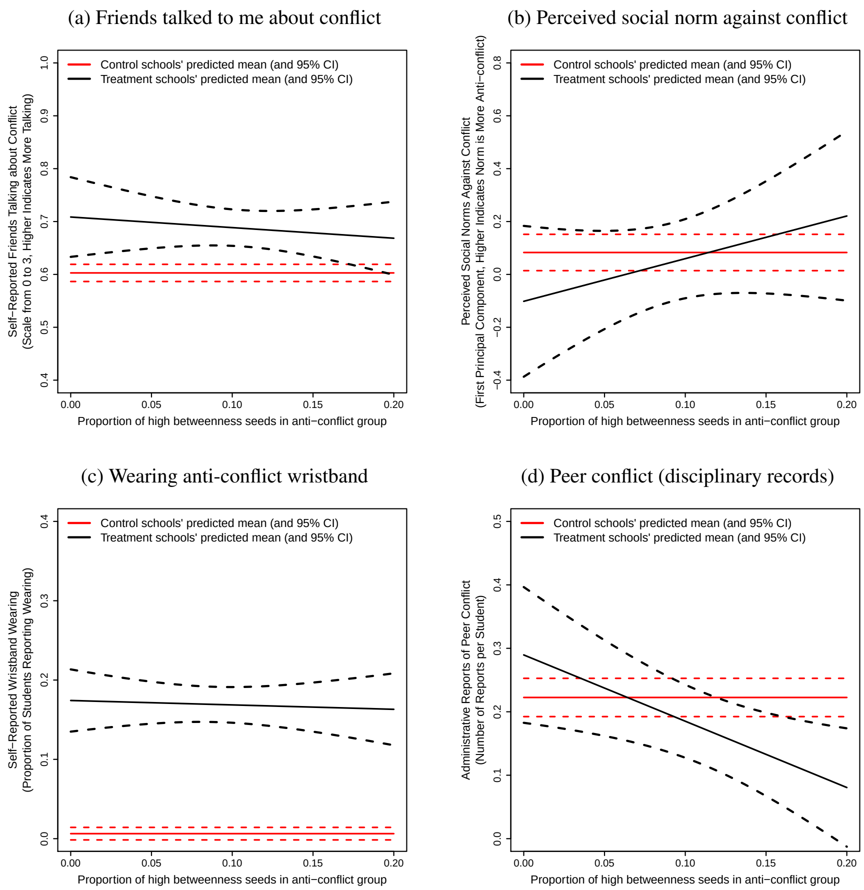

Descriptive norms
Two−tailed p = 0.376
Frequency
Effect Estimates under Null, Observed = 0.069
Descriptive norms (Social referent seeds)
Two−tailed p = 0.013
Frequency
Effect Estimates under Null, Observed = 0.635
Wristband wearing
Two−tailed p = 0.013
Frequency
Effect Estimates under Null, Observed = 0.02
Wristband wearing (Social referent seeds)
Two−tailed p = 0.009
Frequency
Effect Estimates under Null, Observed = 0.073



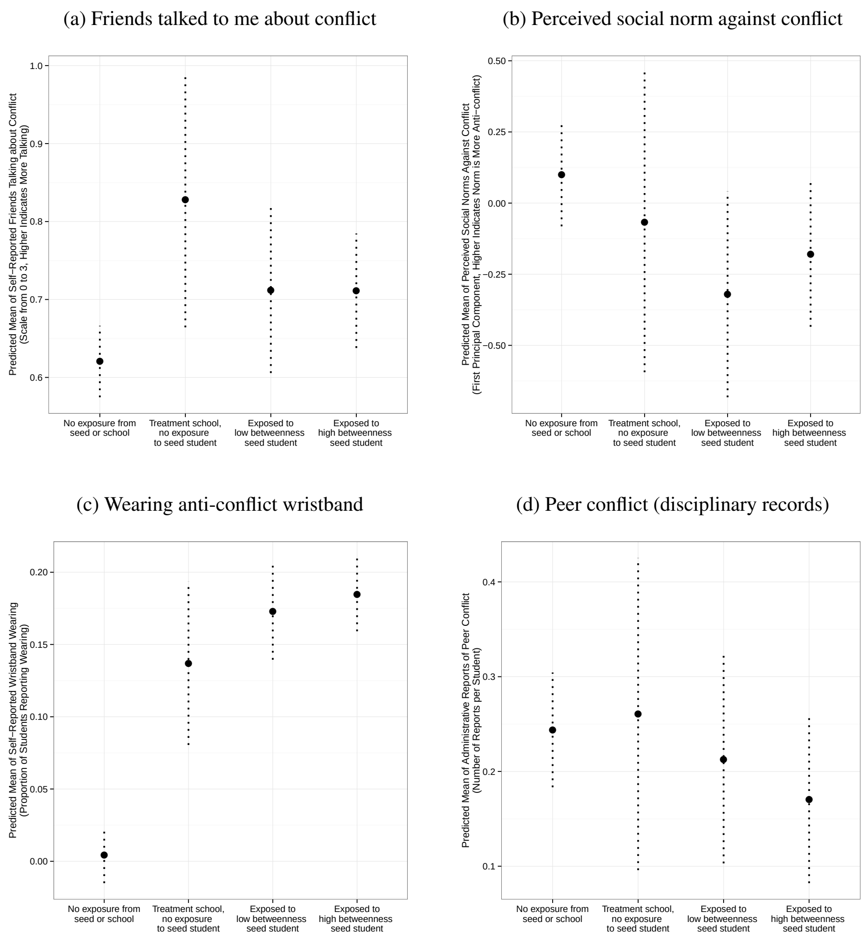


Our results comport with the estimates found in our primary analysis. Overall, we found a statistically significant influence effect on wristband usage ($p < 0.05$). Among the peers of referents, we found statistically significant influence effects on wristband usage ($p < 0.01$) and prescriptive and descriptive norms ($p < 0.05$). Recall that this method does not allow for assessing between-school variation, and thus our comparisons with this methodology are restricted to social influence effects and do not incorporate any differences attributable to the climate of the school.

Fig. S6. Re-analysis of Fig. 2 and 3 using betweenness in place of indegree



Average school-level climate effects caused by anti-conflict groups, interacted with proportion of high betweenness seeds in the group. Lines represent estimates of the conditional average potential outcome with respect to the proportion of high betweenness seeds, and are surrounded by 95% confidence intervals estimated using heteroskedasticity-robust standard errors for the regression fit.



Causal social influence effects from seed students (betweenness in place of indegree). The figures illustrate estimates of average potential outcomes and 95% confidence intervals generated via randomly permuting treatment assignment under a maintained hypothesis of constant effects. We find no evidence of peer social influence on discipline for peer conflict.

3 Tables and legends

Descriptive statistics for schools

Table S1. Summary statistics for each school block and for overall sample

School Block	% White (self-ID)	Home language	Friends like house	School size	School size in 2011	5th grade size	6th grade size	7th grade size	Network cluster coeff.	Network density	Limited Eng. Prof.	% White	% Black	% Hispanic	% Free reduced lunch
1	0.73	0.42	0.57	368.25	377.00	93.25	87.25	93.75	0.15	0.02	7.25	0.72	0.03	0.19	0.17
	0.11	0.14	0.04	72.86	91.16	14.08	21.65	23.56	0.02	0.00	6.42	0.12	0.01	0.11	0.07
2	0.75	0.35	0.61	389.75	408.50	0.00	129.50	124.50	0.13	0.02	8.25	0.75	0.02	0.11	0.08
	0.11	0.12	0.03	33.64	24.44	0.00	11.15	10.59	0.02	0.00	11.50	0.11	0.01	0.06	0.05
3	0.59	0.44	0.59	226.50	237.50	24.25	66.75	69.75	0.21	0.03	10.75	0.57	0.07	0.30	0.31
	0.16	0.13	0.04	37.71	35.77	24.38	19.55	21.05	0.05	0.01	12.38	0.21	0.03	0.18	0.17
4	0.82	0.20	0.60	715.00	732.50	0.00	239.00	243.75	0.08	0.01	8.00	0.81	0.07	0.08	0.15
	0.04	0.02	0.02	71.98	94.25	0.00	29.25	23.90	0.01	0.00	12.19	0.07	0.04	0.03	0.08
5	0.65	0.41	0.56	122.00	114.50	0.00	20.50	49.75	0.27	0.06	7.25	0.64	0.03	0.27	0.25
	0.33	0.30	0.06	14.90	8.90	0.00	20.61	12.89	0.04	0.01	6.18	0.30	0.02	0.29	0.19
6	0.82	0.23	0.59	528.50	578.25	0.00	43.75	254.00	0.08	0.01	1.75	0.86	0.02	0.09	0.13
	0.10	0.12	0.03	55.98	77.62	0.00	75.78	54.95	0.03	0.00	1.79	0.07	0.01	0.08	0.07
7	0.35	0.71	0.57	335.50	336.00	28.50	104.00	105.25	0.13	0.02	20.75	0.33	0.12	0.47	0.64
	0.17	0.11	0.02	113.25	124.63	28.54	44.43	49.11	0.06	0.01	20.93	0.19	0.02	0.17	0.20
8	0.75	0.35	0.58	492.75	513.75	182.00	205.75	50.25	0.11	0.02	9.00	0.77	0.03	0.08	0.09
	0.12	0.14	0.01	121.78	157.87	80.53	81.80	60.05	0.05	0.01	11.64	0.11	0.01	0.01	0.04
9	0.20	0.84	0.58	320.25	327.25	24.00	72.50	112.50	0.13	0.03	46.00	0.18	0.02	0.74	0.46
	0.07	0.09	0.02	83.26	65.86	41.57	48.57	32.56	0.05	0.01	26.27	0.10	0.03	0.16	0.08
10	0.88	0.11	0.61	318.50	332.25	0.00	50.25	134.00	0.15	0.03	8.25	0.87	0.04	0.06	0.12
	0.05	0.03	0.05	130.99	140.09	0.00	54.41	63.50	0.06	0.01	7.89	0.04	0.03	0.02	0.02
11	0.78	0.31	0.59	512.75	503.25	0.00	71.00	209.75	0.07	0.02	3.75	0.77	0.05	0.12	0.15
	0.11	0.16	0.02	98.68	157.52	0.00	75.67	74.86	0.02	0.00	3.27	0.09	0.04	0.07	0.12
12	0.76	0.28	0.59	468.75	407.25	63.75	132.75	140.00	0.07	0.02	2.75	0.77	0.04	0.06	0.10
	0.12	0.08	0.04	255.36	253.61	77.98	58.46	79.58	0.04	0.01	2.77	0.13	0.03	0.03	0.09
13	0.59	0.41	0.56	481.25	408.00	61.25	151.00	154.25	0.13	0.02	51.25	0.57	0.08	0.28	0.20
	0.25	0.28	0.02	186.18	126.94	69.85	65.42	68.54	0.04	0.01	84.74	0.27	0.09	0.31	0.21
14	0.11	0.49	0.48	403.25	350.00	27.75	86.75	96.50	0.11	0.03	8.25	0.06	0.68	0.23	0.61
	0.03	0.09	0.03	181.59	150.54	48.06	93.28	75.69	0.07	0.02	6.02	0.05	0.18	0.13	0.16
Total	0.63	0.40	0.58	405.93	401.86	36.05	104.34	131.29	0.13	0.02	13.80	0.62	0.09	0.22	0.25
Sample	0.28	0.24	0.05	185.99	190.58	63.96	81.90	81.73	0.07	0.01	29.49	0.29	0.17	0.24	0.22

NOTE: Each block contains four schools; statistics are presented by block to ensure confidentiality of participating schools. “White self-ID” represents the average proportion of students in the blocked set of schools who self-identified as white in the pre-randomization student surveys; “home language” represents the average proportion of students who reported that they speak only English (1) vs. other languages (0) at home in the pre-randomization student surveys; “Friends like house” indicates the average proportion of students who indicated that their friends say that their “house is nice” in the pre-randomization student surveys (1, vs. 0: a measure of peer-perceived socioeconomic status); “school size” is the average school population as measured by the number of students who took the pre-randomization student surveys; the next four variables summarize data obtained from the New Jersey State Department of Education indicating the average total school and the 5th, 6th, and 7th grade population during the year prior to the study (2011); the average network cluster coefficient (range: .02 -.33) and network density (range: .01 - .08) for schools in each block was calculated from student network data gathered in the pre-randomization student surveys; the remaining five variables summarize data obtained from the New Jersey State Department of Education based on 2011 school reporting to the state. We also blocked on latitude and longitude of each school but do not report those statistics in order to preserve school anonymity.

Table S2. Pre-Treatment Self-Reported Characteristics of Eligible and Ineligible Social and Non-Social Referents

Pre-Treatment Self-Reported Variables	**Eligible Social Referents**	**Eligible Non-Social Referents**	**Ineligible Social Referents**	**Ineligible Non-Social Referents**	**Paired *t*-test *p***
Gender	0.41 (0.03)	0.51 (0.01)	0.56 (0.01)	0.50 (0.00)	0.000
Age	12.53 (0.08)	12.52 (0.07)	12.67 (0.06)	12.46 (0.07)	0.000
GPA	3.21 (0.05)	3.12 (0.04)	3.13 (0.04)	3.11 (0.04)	0.192
Has college plans	0.92 (0.02)	0.83 (0.01)	0.91 (0.01)	0.85 (0.01)	0.000
Home language is English	0.73 (0.04)	0.59 (0.03)	0.71 (0.03)	0.62 (0.03)	0.000
Friends say my house is really nice	0.66 (0.03)	0.55 (0.01)	0.68 (0.01)	0.57 (0.01)	0.000
Friends come over every week	0.68 (0.03)	0.48 (0.02)	0.74 (0.02)	0.52 (0.01)	0.000
Mother has college education	0.70 (0.03)	0.72 (0.02)	0.73 (0.02)	0.74 (0.02)	0.787
Moved in last few years	0.13 (0.02)	0.21 (0.01)	0.12 (0.01)	0.19 (0.01)	0.000
Attend sports at schools	0.44 (0.04)	0.32 (0.01)	0.53 (0.02)	0.33 (0.01)	0.000
Date people at school	0.34 (0.03)	0.15 (0.01)	0.44 (0.02)	0.20 (0.01)	0.000
Read books for fun	0.26 (0.03)	0.36 (0.01)	0.21 (0.01)	0.35 (0.01)	0.000
Use Facebook	0.56 (0.04)	0.47 (0.02)	0.60 (0.02)	0.47 (0.02)	0.000
Use MySpace	0.03 (0.01)	0.04 (0.01)	0.04 (0.01)	0.04 (0.00)	0.491
Use Twitter	0.28 (0.03)	0.21 (0.01)	0.35 (0.02)	0.24 (0.01)	0.000
Use Tumblr	0.16 (0.02)	0.13 (0.01)	0.20 (0.01)	0.13 (0.01)	0.000
Use Instagram	0.58 (0.03)	0.35 (0.01)	0.64 (0.02)	0.39 (0.01)	0.000
Positive experiences at school (0-4)	2.50 (0.06)	2.03 (0.03)	2.46 (0.03)	1.99 (0.03)	0.000
Negative experiences at school (0-9)	0.73 (0.07)	0.78 (0.03)	0.86 (0.04)	0.71 (0.02)	0.000
Sports outside of schools	0.84 (0.02)	0.65 (0.02)	0.85 (0.01)	0.69 (0.01)	0.000
Do theater/drama	0.10 (0.02)	0.10 (0.01)	0.09 (0.01)	0.11 (0.01)	0.004
Do music	0.34 (0.03)	0.37 (0.01)	0.30 (0.01)	0.36 (0.01)	0.000
Do lots of homework	0.47 (0.03)	0.44 (0.01)	0.42 (0.02)	0.43 (0.01)	0.744
Has younger siblings	0.51 (0.03)	0.55 (0.01)	0.51 (0.01)	0.55 (0.00)	0.000
Has older siblings	0.70 (0.02)	0.62 (0.01)	0.69 (0.01)	0.61 (0.01)	0.000
Has no siblings	0.05 (0.01)	0.07 (0.01)	0.06 (0.01)	0.07 (0.00)	0.062
White	0.70 (0.05)	0.63 (0.04)	0.71 (0.04)	0.66 (0.04)	0.000
Black	0.09 (0.03)	0.10 (0.02)	0.09 (0.02)	0.10 (0.02)	0.636
Hispanic	0.20 (0.04)	0.22 (0.03)	0.22 (0.03)	0.20 (0.03)	0.203
Asian American	0.04 (0.01)	0.07 (0.01)	0.03 (0.01)	0.06 (0.01)	0.000
Caribbean	0.02 (0.01)	0.02 (0.01)	0.02 (0.00)	0.02 (0.01)	0.048
South Asian	0.03 (0.01)	0.02 (0.00)	0.01 (0.00)	0.02 (0.00)	0.001
Other	0.08 (0.02)	0.11 (0.01)	0.07 (0.01)	0.10 (0.01)	0.000
Middle East	0.01 (0.00)	0.02 (0.00)	0.00 (0.00)	0.01 (0.00)	0.000
Native American	0.00 (0.00)	0.00 (0.00)	0.01 (0.00)	0.00 (0.00)	0.996

Note: Heteroskedasticity-robust standard errors clustered by school in parentheses. Mean estimates include adjustment for block fixed effects. Paired t-test p refers to the p-value from a t-test comparing each variable's school-level means for social referents against the school-level means for non-social referents, pooled across eligible and non-eligible students.

Table S3. Factor loadings for Prescriptive norm items (behaviors perceived to be desirable at school)

Survey Item	Wave 1	Wave 2
Item 1	-0.29	-0.27
Item 2	-0.26	-0.26
Item 3	0.32	0.26
Item 4	0.30	0.27
Item 5	0.28	0.23
Item 6	0.27	0.25
Item 7	-0.29	-0.28
Item 9	0.09	0.10
Item 10	-0.36	-0.34
Item 11	-0.34	-0.33
Item 12	-0.40	-0.38
Item 13		-0.38

Table S4. Factor loadings for Descriptive norm items (behaviors that are typical at school)

Survey Item	Wave 1	Wave 2
Item 1	-0.38	-0.34
Item 2	-0.39	-0.34
Item 3	-0.32	-0.31
Item 4	-0.40	-0.37
Item 5	-0.25	-0.27
Item 6	-0.36	-0.32
Item 7	-0.28	-0.31
Item 8	-0.37	-0.36
Item 9	-0.07	-0.05
Item 10	0.04	0.10
Item 11	0.15	0.14
Item 12	0.07	0.07
Item 14		-0.32

Table S5. Causal social influence effects from seed students.

	Friends talk about conflict	Prescriptive norms	Wristband	Peer conflict
Exposed to social referent seed student	0.747	0.007	0.232	0.170
	(0.032)	(0.113)	(0.012)	(0.037)
Exposed to non-referent seed student	0.696	-0.468	0.124	0.190
	(0.051)	(0.163)	(0.016)	(0.052)
Treatment school, no exposure to seed student	0.695	-0.825	0.095	0.152
	(0.084)	(0.254)	(0.025)	(0.072)
No exposure to seed or school	0.624	0.141	0.002	0.200
	(0.020)	(0.075)	(0.008)	(0.027)

Estimates of average potential outcomes using covariate-adjusted inverse probability weighting estimator detailed in the Supplementary Information section *Analysis of Social Influence Effects*, computed on the subpopulation of 2,451 students who had a positive probability of falling into all four levels of exposure. Standard errors (in parentheses) are computed under a maintained hypothesis of constant effects. Contrasts represent causal differences in outcomes attributable to different configurations of treated peers. A graphical version of this table is presented in Figure 4 in the main text of the paper.

Table S6. Average school-level *climate* effects caused by anti-conflict groups.

	Friends talk about conflict	Prescriptive norms	Wristband	Peer conflict
School level main effect	0.09	-0.01	0.16	-0.06
Heteroskedasticity Robust SE	(0.02)	(0.08)	(0.01)	(0.03)
Bayesian Bootstrap SE	(0.01)	(0.05)	(0.01)	(0.03)
Wild Bootstrap SE	(0.02)	(0.08)	(0.01)	(0.03)

Regression-based estimates of mean differences between treatment and control schools. Estimates computed using ordinary least squares with covariate controls and fixed effects for randomization block. Standard errors are reported in parentheses. Bayesian bootstrap standard errors generated using 5,000 replicates, applying weights to each school drawn from the standard uniform distribution. Wild bootstrap standard errors generated using 5,000 replicates, multiplying residuals by draws from the standard normal distribution.

Table S7. Average school-level *climate* effects caused by anti-conflict groups, interacted with proportion of social referent seeds in the group.

	Friends talk about conflict	Prescriptive norms	Wristband	Peer conflict
School effect with 0%				
Social Referents	0.08	0.05	0.15	0.04
Heteroskedasticity Robust SE	(0.03)	(0.15)	(0.02)	(0.07)
Bayesian Bootstrap SE	(0.03)	(0.11)	(0.02)	(0.05)
Wild Bootstrap SE	(0.03)	(0.14)	(0.02)	(0.07)
Interaction: Proportion of				
Referents in Treatment group	0.10	-0.53	0.12	-0.85
Heteroskedasticity Robust SE	(0.29)	(1.06)	(0.18)	(0.46)
Bayesian Bootstrap SE	(0.22)	(0.80)	(0.13)	(0.35)
Wild Bootstrap SE	(0.30)	(1.07)	(0.18)	(0.47)

Regression-based estimates of the effects of treatment conditional on the proportion of social referent seeds in the group, under a linear approximation. "School effect with 0% Social Referents" refers to the estimated value of treatment when the group contains no Social Referents. "Interaction: Proportion of Referents in Treatment groups" refers to the first derivative of the effect size with respect to the percentage of the group that were referents. Estimates computed using ordinary least squares with covariate controls and fixed effects for randomization block. A control is also included for the expected value of the seed group composition in the school (over all possible randomizations). Standard errors are reported in parentheses. Bayesian bootstrap standard errors generated using 5,000 replicates, applying weights to each school drawn from the standard uniform distribution. Wild bootstrap standard errors generated using 5,000 replicates, multiplying residuals by draws from the standard normal distribution. A graphical version of this table is presented in Figure 2 in the main text of the paper.

4 Other

Link to Pre-registration analysis plan:

Pre-registration analysis plan

A Dataverse DOI link to all replication code will be made public upon publication; data are available from Princeton University archive, with appropriate IRB approvals, due to confidentiality of school student data.

Further Acknowledgements:

We received research assistance from Ellen Alpert, Jonathon Baron, Courtney Bearns, Kassandra Birchler, Jose Drost-Lopez, Izzy Gainsburg, Robin Gomila, Tamara Halperin, Monica Hannush, Alexandra Lieberman, Molly Offer-Westort, Alan Paluck, Maitane Sesma, Rebecca Shaw, Pavita Singh, Margaret Tankard, and Aaron Yin. We acknowledge the feedback of Donald Green, Winston Lin, Deborah Prentice, Cyrus Samii, Mario Small, and of participants in the Paluck lab and in seminars at CIFAR, EGAP, Yale and Rutgers University.