

HOW DOES PEER PRESSURE AFFECT EDUCATIONAL INVESTMENTS?*

LEONARDO BURSZTYN AND ROBERT JENSEN

When effort is observable to peers, students may try to avoid social penalties by conforming to prevailing norms. To test this hypothesis, we first consider a natural experiment that introduced a performance leaderboard into computer-based high school courses. The result was a 24 percent performance decline. The decline appears to be driven by a desire to avoid the leaderboard; top performing students prior to the change, those most at risk of appearing on the leaderboard, had a 40 percent performance decline, while poor performing students improved slightly. We next consider a field experiment that offered students complimentary access to an online SAT preparatory course. Sign-up forms differed randomly across students only in whether they said the decision would be kept private from classmates. In nonhonors classes, sign-up was 11 percentage points lower when decisions were public rather than private. Honors class sign-up was unaffected. For students taking honors and nonhonors classes, the response depended on which peers they were with at the time of the offer, and thus to whom their decision would be revealed. When offered the course in a nonhonors class (where peer sign-up rates are low), they were 15 percentage points less likely to sign up if the decision was public. But when offered the course in an honors class (where peer sign-up rates are high), they were 8 percentage points more likely to sign up if the decision was public. Thus, students are highly responsive to their peers are the prevailing norm when they make decisions. *JEL* Code: I21.

I. INTRODUCTION

It has long been suggested that students may be motivated as much by the desire to gain social approval (e.g., being popular or fitting in) or avoid social sanctions (e.g., being teased, made fun of or ostracized) as they are by the future benefits of education (Coleman 1961). An important question then arises as to whether and how student effort or investments are affected by such peer

*We thank Nava Ashraf, Ernesto Dal Bó, Leigh Linden, Aprajit Mahajan, Torsten Persson, Bruce Sacerdote, Noam Yuchtman, numerous seminar participants, and Lawrence Katz, Andrei Shleifer, and four anonymous referees for comments and suggestions. We thank Pedro Aratanha, Andrea Di Miceli, Stefano Fiorin, Craig Jones, Vasily Korovkin, Matthew Miller, and Benjamin Smith for excellent research assistance. We are grateful to the UCLA Anderson Price Center and the California Center for Population Research for financial support. Our study was approved by the UCLA Institutional Review Board and the Los Angeles Unified School District Committee on External Research Review.

© The Author(s) 2015. Published by Oxford University Press, on behalf of President and Fellows of Harvard College. All rights reserved. For Permissions, please email: journals.permissions@oup.com

The Quarterly Journal of Economics (2015), 1329–1367. doi:10.1093/qje/qjv021.

Advance Access publication on May 14, 2015.

social concerns or peer pressure.¹ Are students willing to deviate from what they privately believe to be the optimal scholastic effort or investment decision just because of such social concerns? In this article, we test this hypothesis using both a natural experiment and a field experiment.

Despite the perception that peer pressure is widespread, there is very little direct empirical evidence of its effects.² Some studies have found peer social concerns in the workplace. For example, Mas and Moretti (2009) find that the productivity of supermarket cashiers is affected by coworkers who can see them (particularly those they interact with more), but not those who can't. Bandiera, Barankay and Rasul (2010) find that the productivity of fruit pickers is affected by those working alongside them, but only when they are friends. Adolescence is believed to be the period of greatest vulnerability to peer pressure, during which the desire to be popular or fit in is felt most acutely (Brown 2004). Adolescents may be more likely to give in to such pressure and engage in behaviors that can have long-term effects. Although many studies have found peer effects in education, there are many mechanisms through which they might occur.³

We begin by examining how the introduction of a system that revealed top classroom performers affected subsequent performance. The natural experiment we consider was applied to a computer-based learning system used in over 100 high schools located predominantly in one U.S. state. The system is primarily used for remedial English and math courses, particularly in preparation for a high-stakes, statewide high school exit exam. Prior

1. We define peer pressure as students taking actions that deviate from what they privately consider to be the optimal action (what they would do if others would not observe their actions) to achieve social gains or avoid social costs. Peer pressure thus doesn't just refer to active efforts by peers to persuade others, but could also include passive effects such as not undertaking an action for fear of peer social sanctions.

2. Some studies in social psychology measure peer pressure through direct survey questions, such as by asking whether a student has faced pressure from others to undertake certain actions (Brown 1982; Brown, Clasen, and Eicher 1986; Santor, Messervey, and Kusumakar 2000). However, there is some concern with using such subjective self-reports, and it is difficult to link these responses directly and causally to specific behaviors.

3. Sacerdote (2001), Zimmerman (2003), Carrell, Fullerton, and West (2009), Duflo, Dupas, and Kremer (2011) and Carrel, Sacerdote, and West (2013). See Epple and Romano (2011) for a summary. Bursztyn et al. (2014) and Cai, De Janvry, and Sadoulet (2012) examine channels of peer influence in financial settings.

to the change, students would answer multiple-choice questions and receive private feedback on whether their answers were correct. One month into the 2011–12 school year, without any advanced notice or explanation, the system was changed. Students were now awarded points for correct answers. Simultaneously, home screens provided tabs revealing the names of the top three scorers in the classroom, the school, and among all users of the system, as measured by cumulative points received for the past week, month, and all time. Finally, each tab showed students their own rank (in the classroom, school, and among all users, for the past week, month, and all time). There were no other changes to the system.

We find that the introduction of the leaderboard led to a 24 percent decline in performance. We also provide evidence that these results are driven by an aversion to being on the leaderboard. Because students had already been using the system for over a month before the change, they would have had some private information about their own performance, and thus their risk of showing up on the leaderboard if they continued performing at their previous level. We find that students in the top quartile of the prechange distribution of performance, those most at risk of showing up on the leaderboard, on average had a 40 percent decline in performance. This decline comes primarily through attempting fewer questions (not getting fewer questions correct), and includes reduced discretionary use of the system outside of school; both are consistent with an active choice to reduce effort. These students cut back almost immediately, with declines evident on the very first day of the change. By contrast, students at the bottom of the prechange performance distribution did slightly better following the change. The pattern across the distribution of prechange performance is monotonic; on average, the better you were performing before the leaderboard was in place (and thus the greater the risk of being in the top three), the more your performance declined afterward. In other words, it appears that at least some students were willing to work hard and do well only as long as their classmates wouldn't know about it.

To further isolate and test for peer pressure, and to see whether these effects apply beyond remedial students, we turn to our field experiment. In four low-income Los Angeles high schools, we offered eleventh-grade students complimentary access to an online SAT preparatory course from a well-known

test preparation company. Across students within classrooms, we randomly varied whether the sign-up forms emphasized that the decision to enroll would be kept private from the other students in the classroom. In particular, students were either told that their decision to enroll would be kept completely private from everyone *including* the other students in the room, or *except* those students. Notably, the sole difference between sign-up forms in our “private” and “public” treatments was a single word (*including* versus *except*).

We use both honors and nonhonors classes for the experiment. The prep class is an educational investment, and making it observable to peers could carry different social costs in settings where the norms on the acceptability of effort differ, such as in the two types of classes. We find that observability has a large impact on sign-up rates, and the effects differ across settings. In nonhonors classes, sign-up is 11 percentage points lower when students believe others in the class will know whether they signed up, compared with when they believed it would be kept private. In honors classes, there is no difference in sign-up rates under the two conditions.

Consistent with these results being driven by peer social concerns, in nonhonors classes, students who say it is important to be popular are less likely to sign up when the decision is public rather than private, whereas students who say it is not important are not affected at all. In honors classes, students who say that it is important to be popular are slightly more likely to sign up when the decision is public, whereas those who say it is not important are again unaffected. In both cases, students concerned with popularity move in the direction of the locally prevailing norm when the decision is public, whereas those unconcerned with popularity are unaffected.

The differential response to observability by class type could still be consistent with explanations other than peer social concerns. For example, students in honors and nonhonors classes may differ from each other in many ways, which may affect how much they care about privacy or how they respond when decisions are observable. Though this would not change the policy implication that observability affects sign-up in nonhonors classes, to test peer pressure even more cleanly we can address this selection problem and make the set of students we examine in honors and nonhonors classes more similar by focusing on students taking both types of classes. Students are free to choose

whether to take an honors or nonhonors class for every subject for which both are available. To fix ideas, consider the set of students taking exactly two honors classes (hereafter “two-honors” students). Honors classes are spread throughout the day, but our team showed up for just two periods. The timing of our arrival will find some two-honors students in an honors class and others in a nonhonors class. Just as important, the timing of our visit, and therefore which type of class we find them in, will be uncorrelated with student characteristics. Thus, although this approach does not explicitly randomize peers, the set of two-honors students who happen to be sitting in one of their honors classes when we arrive and conduct our experiment should be similar in expectation to those who happen to be sitting in one of their nonhonors classes—all that will differ is whether they are at that moment sitting with their honors or nonhonors peers. Because we are not actually changing a student’s peers at all (or their teachers, schools, neighborhoods, or anything else in their environment), we can rule out most other channels through which peers may influence each other. We capture the effect of varying just to which of a student’s peers the sign-up decision will be revealed, and thus whether and how those peers punish or reward observable effort.

Examining the set of all students taking just some honors classes (“some-honors” students), we find that making the decision to enroll public rather than private decreases sign-up rates by 15 percentage points when they are in one of their nonhonors classes (where sign-up among their classmates is low). By contrast, making the decision public *increases* sign-up rates by 8 percentage points when they are in one of their honors classes (where sign-up rate among classmates is high). Viewed another way, when decisions are observable to peers, the sign-up rate is over 20 percentage points lower when they are with their nonhonors peers rather than their honors peers.

Both settings suggest peer social concerns affect educational investment and effort. When faced with the trade-off between the future benefits of academic effort and the present potential social costs, some students choose to reduce effort and performance (though conforming to the locally prevailing norm can also induce greater effort or investment, as observed in honors classes under the public treatment). The fact that we find similar effects in two entirely independent settings, plus the fact that many investments or efforts students may make are observable to peers

(e.g., raising a hand in class, seeking extra help, or earning academic honors), suggests that such effects may be found more widely. These results are also relevant to the literature examining the role of schools and neighborhoods in the educational outcomes of poor and minority students (Oreopoulos 2003; Jacob 2004; Kling, Liebman, and Katz 2007; Dobbie and Fryer 2011; Fryer and Katz 2013). Finally, the results may also carry implications for the design of school policy or practices. Changing either norms or peers is likely to be quite difficult, particularly on a large scale; but for at least some activities, changing the extent to which behavior is observable to peers is likely to be less so.⁴

Because both settings we examine include primarily Hispanic students, our results are generally supportive of the “acting white” hypothesis, whereby minorities may face peer sanctions for certain behaviors (Fordham and Ogbu 1986; Austen-Smith and Fryer 2005; Fryer 2011; Fryer and Torelli 2010).⁵ More broadly, Austen-Smith and Fryer (2005) and Fryer (2007) model the trade-off between group loyalty and economic success.⁶ For example, when students face a tension between investing in activities rewarded by the labor market and signaling loyalty to a group, one equilibrium involves sorting wherein high-ability individuals invest in labor market-oriented activities rather than those likely to increase acceptance by the group, and low-ability individuals choose the reverse. The differential response we observe by class type is consistent with such sorting, with social penalties only for students in nonhonors classes. The fact that we consider an SAT prep course is particularly relevant, given that such an investment signals a likelihood that the individual may leave the group and is thus the type of behavior we expect to be sanctioned under these models. Finally, our results are also relevant to other general models of social

4. Also, the extent to which changing peer groups can help is limited by the fact that if enough students are moved across groups, the dominant norm may change. Carrell, Sacerdote, and West (2013) show that even when you construct peer groups, students may re-sort into more homogeneous subgroups.

5. However, for both the natural and field experiments, we lack sufficient variation in ethnicity or race to test whether similar effects are found for other students or are specific to minorities.

6. These models can account for a variety of social interactions and group dynamics that have been found for several ethnic and cultural groups (e.g., Gans 1962; Ausubel 1977; Lee and Warren 1991; Berman 2000).

interactions (Akerlof 1997; Becker and Murphy 2000; Durlauf and Ioannides 2010; Postlewaite 2011), including those that examine the role of culture in shaping interactions.

The remainder of this article proceeds as follows. In Section II, we discuss the natural experiment in more detail. Section III discusses the field experiment, and Section IV concludes. All appendix material is available in the Online Appendix.

II. LEADERBOARD NATURAL EXPERIMENT

II.A. Background and Policy Change

Many schools use in-class, computer-based learning materials created and operated by private companies. The company responsible for the software we consider was operating in over 200 high schools across several states (though primarily in just one). The sample of schools using the software is not random. These schools have a higher minority share and local poverty rate compared with other schools in the state where the system is most widely used.

The most widely used courses are tenth- and eleventh-grade remedial English and math, including those designed for statewide high school exit exams; we restrict our analysis to these courses. Schools require students to take these courses if they scored in the lowest proficiency levels on the previous year's statewide standardized test or if they failed the high school exit exam. The fact that these students are low-performing is relevant for generalizability, since they might, for example, be more sensitive to peer social stigma than is the average student.

Students are given individual online accounts. When logged in, they have access to a database of questions. Questions are multiple choice, and after each question students receive private feedback on whether their answer is correct. The questions are organized into modules that typically follow in-class instruction. Students have some discretion in how many questions they answer (the database was sufficiently large that students would not run out of questions). Students can also access the system at any time from any computer or device with an Internet connection for additional, voluntary learning opportunities.

On September 20, 2011, without any prior notice or explanation, the company introduced a point system and a series of rankings and leaderboards, intended to encourage and motivate

students. Students were now awarded 1,000 points for answering a question correctly on the first try, 325 points for a correct answer on the second or third tries, and no points after that. There was no penalty for incorrect answers; thus students could increase their score by getting more questions correct on the first, second, or third tries or by attempting more questions.

In addition, students could access a series of tabs showing the names of the top three scorers (based on cumulative points, starting at 0 on the day of the change) in their class, school, and among all users of that course. These leaderboards were updated in real time and were separately available for the past week, month, and all time. Finally, students could also access their own rank (again, in their class, school, and among all users, and for the past week, month, and all time). The system was otherwise unchanged during the period of our analysis. These changes were introduced at the same time for all students and across all schools.

II.B. Data and Empirical Analysis

We have data for the universe of questions answered, with each student uniquely identified by an ID code. However, we have no other data on students besides their first and last names. We exclude classrooms with fewer than five students and those not using the system before the change. This leaves us with a sample of over 5,000 students across more than 100 schools.

Because the data are click-based, if a student does not attempt any questions on a particular day, they have no data for that day. We recode such cases to zero, but only if at least two other students in their class attempted a question on that day (since there are some days where the system is not used at all by the class). The results are robust to other thresholds, or not recoding at all, since most students attempt at least one question on days when others in their class also do.

To explore the possible role of peer social concerns, we exploit the fact that the potential for being newly exposed to such sanctions by the leaderboard will be greater for some students than others. Because students had been using the system about a month before the change, they would have had some information or signal about their own performance (feedback could also have come from exams, exercises, or directly from teachers). This would likely have included some signal of relative performance

as well. For example, in the extreme cases, a student who was getting most answers correct will likely perceive a higher leaderboard risk than a student who almost never got any answers correct. Therefore, students would have had an approximate sense of whether, if they continued their performance unchanged, they were likely to be among the top performers in the class and have this information revealed to others through the leaderboard. These are the students that we predict will be the most likely to reduce effort if a fear of peer sanctions is operative. We therefore construct a measure of leaderboard risk by computing the distribution of prechange performance within each classroom, based on the number of correct answers students had in the month prior to the change.

As preliminary visual evidence, Figure I plots the average number of correct answers per day, separately by quartile of the within-classroom distribution of prechange performance. We also fit linear trends for the pre- and postchange periods separately, along with 95 percent confidence bands. For previously high-performing students (quartiles 3 and 4), performance declines on the day that the leaderboard is introduced (this first-day drop is statistically significant in regressions for quartile 4 if we include just the day before and day of the change). The number of correct answers then stays lower for the remainder of the period (this persistence is confirmed by regressions that exclude the first day or first few days after the change). By contrast, there is little to no decline for students in quartiles 1 and 2 (with perhaps a slight increase for quartile 1).

We can confirm this visual evidence with regressions exploring how performance changed on introduction of the leaderboard, again separately by quartile of prechange performance:

$$Y_{i,t} = \beta_0 + \beta_1 Post_t + \alpha_i + \epsilon_{i,t}, \text{ for quartile} = \{1, 2, 3, 4\},$$

where $Y_{i,t}$ is the number of questions answered correctly by student i on day t , $Post$ is an indicator for before versus after the policy changes, and α_i are student fixed effects. We trim the post period to one month to match the approximately one month that students were using the system prior to leaderboard introduction (the results are robust to using one- or two-week intervals; see Online Appendix Tables A.I and A.II). The results are also robust to including time trends and differential pre- and postchange trends (see Online Appendix Table A.III). We note that comparing differential changes across

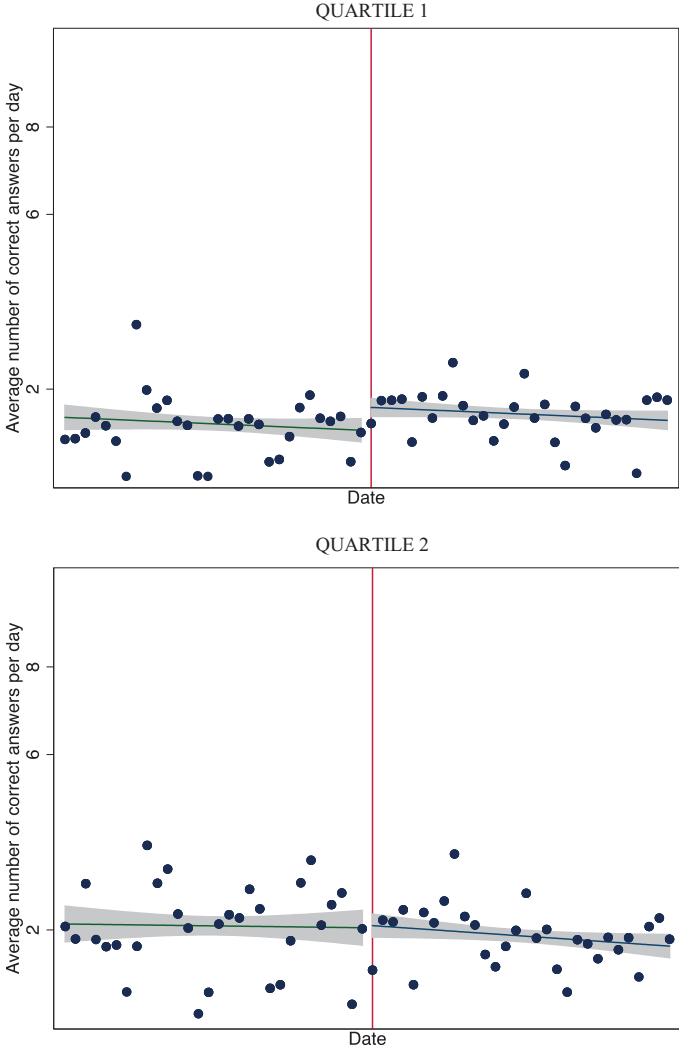


FIGURE I

Average Number of Correct Answers per Day

These figures plot, for each day in the period from 30 days before to 30 days after the introduction of the new system, the average number of correct answers per day. Each figure plots a different quartile of the within-classroom distribution of the total number of correct answers during the month prior to the introduction of the new system. There are 60 observations plotted per quartile. The figures also fit linear trends separately before and after the introduction of the new system, and the 95 percent confidence interval associated with the trends. The vertical line corresponds to the day of the introduction of the new system, September 20, 2011.

(continued)

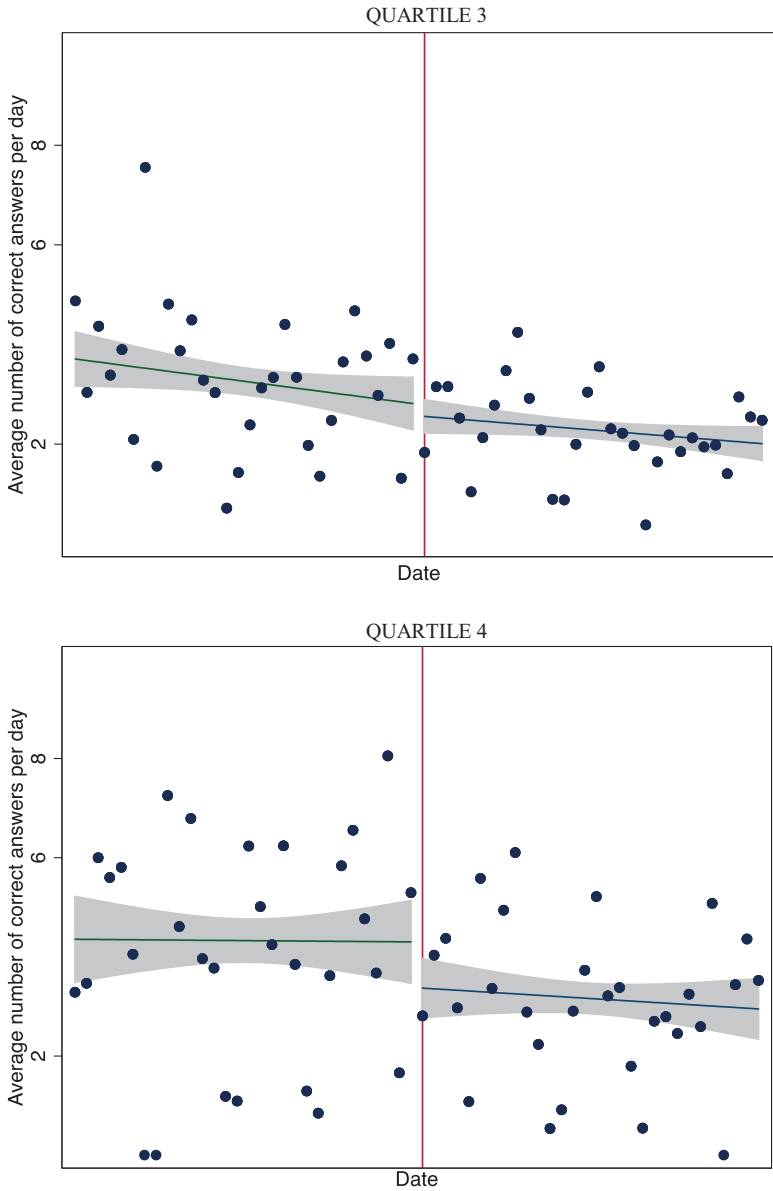


FIGURE I
(continued)

TABLE I
EFFECTS OF THE POINTS AND LEADERBOARD SYSTEM

Dependent variable	(1)	(2)	(3)	(4)	(5)
	Number of correct answers per day				
Post-system change dummy	-0.6266***	0.1717**	-0.5391***	-1.2486***	-1.9340***
	[0.117]	[0.079]	[0.117]	[0.183]	[0.254]
Mean of dependent variable before the change	2.57	1.27	2.43	3.47	4.81
Observations	95,342	37,171	22,978	20,427	14,766
R-squared	0.185	0.161	0.159	0.173	0.174
Sample	Full	Quartile 1	Quartile 2	Quartile 3	Quartile 4

Notes. This table presents OLS regressions of the number of correct answers per day on a dummy on whether the date is after the introduction of the points and leaderboard system. All columns restrict the analysis to the time window between one month before the introduction and one month after it. Column (1) presents the results for the entire sample. Columns (2)–(5) present results by quartile of the within-classroom distribution of the total number of correct answers during the month prior to the introduction of the new system. All regressions include student fixed effects. Robust standard errors clustered by classroom in brackets. *** $p < .01$, ** $p < .05$, * $p < .1$.

quartiles also allows us to net out any changes in other factors that affect all students equally.

Table I provides the results (robust standard errors clustered by classroom in brackets).⁷ Column (1) shows that the effect of the program across all students was negative. After the system is introduced, on average students answer 0.63 fewer questions correctly per day (significant at the 1 percent level). This is a 24 percent decline from the baseline of 2.6. Columns (2)–(5) provide the results for each quartile of prechange performance separately (results are similar if we pool the samples and add interactions). For students in the top quartile in column (5), the change was associated with answering 1.93 fewer questions correctly per day (significant at the 1 percent level). This represents a 40 percent decline from the prechange baseline of 4.8.⁸

7. The imbalance in observations across quartiles arises because most class sizes are not perfectly divisible by 4. The standard way most software creates quartiles results in more observations being assigned to quartile 1 and fewer to quartile 4, as observed here (e.g., for the set of numbers 1,2,3,4,5) the first cut-off will be defined as 2, putting two observations in the first quartile).

8. In separate results (available on request), we find that the effects are similar for boys and girls (we estimate sex by matching first names to the Social Security Administration database of gender frequencies by name).

As we move down from the top quartile in the prechange performance distribution, the effects on performance become less negative, and eventually for the bottom quartile turn positive (in all cases, the results are significant at the five percent level or better. We can also reject equality of the effects for each pairwise combination of quartiles).⁹ Again, these results are suggestive of a role for social sanctions, since it is the students who likely perceive the greatest risk of being in the leaderboard, and thus having their high performance publicly revealed, who cut back the most.¹⁰

So far, our analysis has examined simple before versus after comparisons around the date of the change. To explore the plausibility that the introduction of the points and leaderboard system caused these changes, we can consider the uniqueness of these results. Thus, in a series of placebo tests, we run the same regressions as before but assign the “change date” to every alternate date starting one month after the true change date and ending one month before the end of the school year. For quartile 4 on its own, none of the 218 available placebo dates yield a greater decline in performance than the -1.93 found for the true change date. None yield a point estimate greater than 1.93 either; so even in a two-sided test, no other date in our sample yields as large a change in performance for this previously top-performing group as the day of the leaderboard introduction. Figure II provides a histogram with the distribution of placebo treatment effects, and shows that the estimated decline around the true date is an outlier in terms of sustained changes in student performance around any specific date. Such large and sustained increases or decreases in performance never occur for any other quartile either. Running the placebo tests for the other quartiles, no alternate date yields an estimated increase or decrease in performance of 1.93 .¹¹

9. If we include time trends to capture any general trends in performance over this period, the point estimates are negative for all quartiles, though essentially zero for quartile 1, as shown in Online Appendix Table A.III.

10. If all students equally feared peer social sanctions, we might expect them all to cut back to zero. However, with heterogeneity in disutility from peer sanctions (for example, in our field experiment we show differences in effects by self-rated importance of popularity) then we will shift from a situation where top performers are the highest ability or most motivated to one where they are instead those who care the least about peer stigma (or who are perhaps less likely to actually face such stigma).

11. We also implement an alternative placebo exercise using a permutation test (Conley and Taber 2011). Restricting ourselves to quartile 4 and the original period

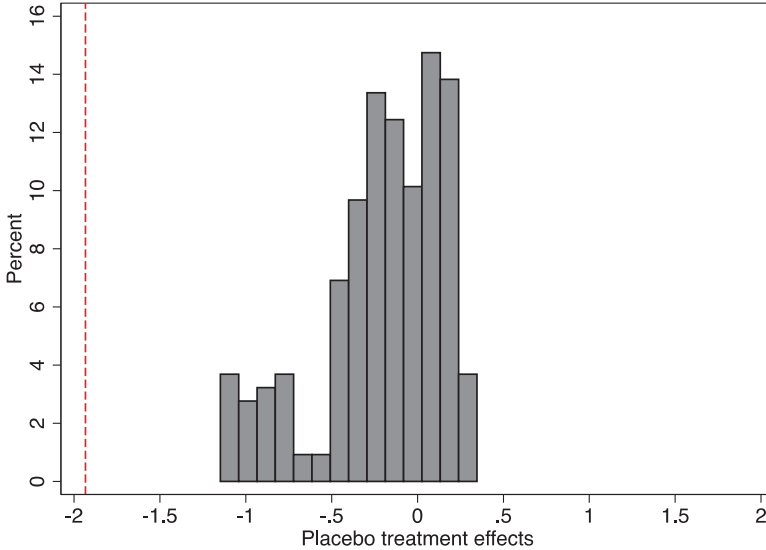


FIGURE II

Distribution of Placebo Treatment Effects for Quartile 4

This histogram displays the distribution of placebo treatment effects estimated for quartile 4 of the within-classroom distribution of the total number of correct answers during the month prior to the introduction of the new system. We run the same regressions as in our main specification, but assign the introduction of the point and leaderboard system to every other month, starting one month after the true date of the change, and ending one month before the end of the school year; there are 218 such days plotted here (each of these regressions has a different number of observations). The dashed line represents our estimated treatment effect for quartile 4 (-1.93).

II.C. Alternative Mechanisms

Though we lack the cleaner experimental design explored below that can help isolate peer pressure as the underlying driving mechanism, we can provide suggestive evidence against several alternative explanations. We first consider whether the leaderboard may have created competitive pressure that adversely affected previously top-performing students. If such

of one month before and after the introduction of the leaderboard, we assign “post-change” status to randomly selected groups of 30 days from that period. We then compare our original estimates to the placebo estimates from 1,000 randomly constructed samples. We never find placebo estimates equal to or larger in absolute value than our estimated treatment effects, so the p -value of this permutation test is .00.

effects were present, we might expect students to perform worse on questions they attempted. However, estimating similar regressions as before, Online Appendix Table A.IV shows that for the full sample, the percent of questions attempted that are answered correctly actually increases slightly after the leaderboard is introduced. For the top quartile, there is a two-point decline in percent correct (from a base of 63 percent); still, over 90 percent of the total decline in questions answered correctly is attributable to attempting fewer questions, which is consistent with an active effort to avoid the leaderboard.¹²

Similarly, if competitive pressure is adversely affecting performance, we might expect less of a decline when the system is used outside of school. Students working outside of school can take as long as they want to answer questions and seek assistance from other people or resources. When working from home they are not facing off against other students in real time. Thus, the system affords them a great deal of opportunity to improve their performance in a less competitive environment if they choose to (points earned at home count toward the leaderboard). However, the bottom panel of Online Appendix Table A.IV shows that performance at home responds the same as use at school. Beyond suggesting competitive pressure is not likely at play, declines in discretionary home use again suggest an active choice to answer fewer questions and avoid the leaderboard.¹³

It is also possible that the information given on rank affected performance (e.g., Barankay 2012; Tran and Zeckhauser 2012; Ashraf, Bandiera, and Lee 2014). For example, the top-performing students may not have known before how well they were doing relative to classmates and may then decide that they don't need to work as hard. However, since the change took place over a month into the school year, students would have already had some information on their performance from exams, assignments, or feedback from teachers. In addition, as noted, the decline for previously top-performing students occurred on the first day of the change. Since all students started

12. This may even be a lower bound, as students trying to avoid the leaderboard may intentionally answer questions incorrectly. Though students may also choose to avoid competition by not answering questions or may be able to answer fewer questions under pressure, so we cannot completely dismiss such pressure.

13. Declines for use at home also suggests that our main results are not driven by distraction effects, with students spending so much time checking the leaderboard that they have less time to answer questions.

at zero, during that first session there would not have been enough time for students to answer a lot of questions correctly, infer from this brief performance that they are (persistently) at the top of the class (beyond what they already knew prior to the change), and still have time to cut back enough (including overcoming their strong performance at the start) that we would see a large net decline on that first day. Similarly, it is unlikely that rank gave teachers enough new information on performance to allow them to redirect attention or resources away from students performing well and toward those needing more help on that very first day. However, we cannot completely rule out that rank plays some role in these effects.

We can rule out that the decline in performance is due to the sudden change or newness of the system. First, the decline was not common to all students, and in fact was found only for previously high-performing students, while worse performing students actually improved. Second, the effects appear to persist beyond the first day's decline (as evidenced by Figure I and by the fact that the regression results are robust to excluding the first day or first few days after the change), whereas over time students should become more familiar with the system and improve.

Finally, mean reversion is unlikely to explain our results. Questions are drawn from the database at random, so there is no explicit design that assigns harder questions to good performers. Statistical or incidental mean reversion is unlikely to explain our results. The prechange quartile is based on over a month of performance, so any randomness or luck is likely to have balanced out. In addition, as noted, the biggest change is in the number of questions attempted, not the percent correct; this likely reflects a conscious choice of effort, whereas a student simply on a lucky (unlucky) streak would likely experience a decline (increase) in the percent answered correctly. Finally, the fact that we find no other changes this large in our placebo test around any other date suggests mean reversion is not found at any other time.

II.D. Summary of the Leaderboard Natural Experiment

Overall, the results so far suggest that students actively reduce effort and performance to avoid appearing on the leaderboard. However, we cannot completely rule out other potential

channels, such as information on rank, nor can we isolate peer pressure as the reason for leaderboard aversion. There remains the question of whether such effects are found more widely or among nonremedial students. To address these questions, we turn to our field experiment.

III. FIELD EXPERIMENT

III.A. Experimental Design

We conducted our experiment in the four largest public high schools in a disadvantaged area of south Los Angeles. We visited each school once, between December 2013 and April 2014. The sample was confined to eleventh-grade students, since this is when many students begin preparing for the SAT. The four schools each have around 3,000 students. In addition to being larger on average, these schools have a higher share of students eligible for free and reduced-price meals (84 percent versus 68 percent) and students of Hispanic ethnicity (96 percent versus 69 percent) compared to the average school in the Los Angeles Unified School District (LAUSD). The median income in the ZIP codes around these four schools is also lower than for all schools in the district (\$39,533 versus \$48,898). We would therefore not want to generalize our results to other schools. However, we do note that these schools account for approximately 7 percent of all high school enrollment in the LAUSD. From a policy perspective, low-performing schools such as these are the ones where it is perhaps most important to understand the barriers to educational investments and performance. Finally, we note that despite these differences, the fraction of seniors in these four schools who take the SAT is the same as for LAUSD as a whole (51 percent).

Within each school, our visits were coordinated with principals and counselors to choose on what day we could visit and during which period(s). These considerations were typically about scheduling logistics for the schools and our research team. During the selected periods, we visited honors and nonhonors classrooms across a range of subjects. Overall, we visited 26 classrooms across the four schools, with a total of 825 students (all of whom participated in the study; we did not contact absent students). Neither students nor teachers were informed in advance about the subject of our visit or that there would be an

intervention related to the SAT (principals were informed in advance, but agreed not to communicate the purpose of our visit ahead of time).

Students in the selected classrooms were offered the opportunity to sign up for free access to a commercial, online SAT preparation course. The course was created by a well-known test prep company that students in these schools are familiar with. The course includes practice exams, a library of prerecorded videos and instructional content, live online class sessions, analysis of individual performance, and test-taking strategy.

Prior to our study, no students in these schools were using the course. The company does not currently offer this software to individuals, instead selling subscriptions to schools, who then make it available to individual students (at a cost of about \$200 per student). None of the schools where we conducted the study had purchased this software prior to our intervention. In a separate follow-up survey at one of our schools (conducted after the intervention), we asked students to estimate the cost of the software; on average, they estimated the value at \$260. Thus, especially for these low income students, this is a valuable offer (perceived and actual) that they would be forgoing if they chose not to sign up (confirmed by the fact that sign-up rates are very high when the decision is private). Finding that observability alone is sufficient to deter sign-up would be an indication that these peer social concerns may be powerful.¹⁴

After a brief introduction, students were given a form offering them the opportunity to sign up for the course. In particular, after asking students their name, sex and favorite subject in school, the form contained the following statement: “[Company Name] is offering a free online test preparation course for the SAT that is intended to improve your chances of being accepted and receiving financial aid at a college you like.” The forms then had one of the following two options: “Your decision to sign up for the course will be kept completely private from everyone, except the other students in the room,” which we refer to as the “public” sign-up condition; or “Your decision to sign up for the course will be kept completely private from everyone, including the

14. Though of course not all students plan to take the SAT, and it would be of little value to such students (unless they gave away or sold their online access to someone else, which we did not explicitly preclude).

other students in the room,” which we refer to as the “private” sign-up condition.

Thus, the sole difference between the two forms, shown in Figure III, was a single word: “except” or “including” (in practice, we did not reveal any sign-up decisions). We also note that the difference in expected privacy is for classmates, not teachers, administrators, or parents.

Students were not given any additional information and were told that all questions should be held until after all forms had been collected. When all students had completed the first form, the research team collected them and handed out a second form with additional questions (discussed below; this form can be found in Online Appendix Figure A.I). When students had completed the second form, the research team collected it and handed out assent and consent forms for authorization to access GPA information. The entire intervention took less than 10 minutes.

The forms with the different privacy assurances were pre-sorted in an alternating pattern and handed out to students consecutively in their seats. By randomizing at the level of the student within the classroom, we ensure that students in the public and private sign-up groups were otherwise treated the same in every other way. So for example there are no differences in how the experimenters or teachers treated students with different privacy statements, no differences in encouragement to enroll or in classroom environment or characteristics. We also did not allow students to communicate with each other until all forms were returned, so there would be no contamination across groups and students would not realize that they were being given different terms of privacy (and even if they looked at each other’s desks, as noted, the forms looked identical at a glance because they only differed by one word).

Because the difference between the two forms was just a single word, the treatment was very small and subtle. This makes it less likely that students would respond to the difference, and we therefore likely underestimate the effects of peer pressure. We chose not to implement treatments that made sign-up even more explicitly public, such as by asking students to raise their hands in the class, come to the front of the room, or put their name on a public sign-up sheet in the room. First, doing so would have required a much greater number of classrooms and schools, and thus significantly higher cost, to have reasonable statistical power since treatments like this could only be implemented at the

A. "Public" Signup Sheet

Student Questionnaire

First name: _____

Last name: _____

Gender (please circle one): Female / Male

What is your favorite subject in school? (Please circle one)
 a. Math b. English Language Arts c. History/Social Studies d. PE/Elective

[Company name] is offering a free online test preparation course for the SAT that is intended to improve your chances of being accepted and receiving financial aid at a college you like.

Your decision to sign up for the course will be kept completely private from everyone, except the other students in the room.

Would you like to sign up for the free [Company name] course? (Please pick one option) Yes / No

If yes, please provide the following contact information:

Email address: _____

Phone number: (____) _____

TURN OVER FORM AND WAIT PATIENTLY

Form A337

B. "Private" Signup Sheet

Student Questionnaire

First name: _____

Last name: _____

Gender (please circle one): Female / Male

What is your favorite subject in school? (Please circle one)
 a. Math b. English Language Arts c. History/Social Studies d. PE/Elective

[Company name] is offering a free online test preparation course for the SAT that is intended to improve your chances of being accepted and receiving financial aid at a college you like.

Your decision to sign up for the course will be kept completely private from everyone, including the other students in the room.

Would you like to sign up for the free [Company name] course? (Please pick one option) Yes / No

If yes, please provide the following contact information:

Email address: _____

Phone number: (____) _____

TURN OVER FORM AND WAIT PATIENTLY

Form A338

FIGURE III
Sign-up Sheets

classroom level. Related, introducing variation at the classroom level could introduce more possible random variation in student, classroom or teacher attributes (or implementation of the treatment) across treatment groups that could separately influence sign-up. A second reason is that the method of signing up (i.e., having the public treatment involve raising a hand or staying after class to sign up and the private treatment involve signing up on an individual sheet of paper) could itself affect sign-up rates. Having all students sign up through the same process but varying only a single word for the two groups makes it clearer that it was just the public versus private nature of the decision that explains any difference in sign-up rates. Finally, having a more public sign-up treatment could have led to other forms of peer effects that we want to exclude, such as social learning or coordination.

As noted already, our priors (aided and confirmed by initial pilot testing) were that the social acceptability of undertaking effort or an investment could vary across settings, particularly with respect to academic performance or baseline levels of effort or investment. Therefore, we explicitly chose both honors and nonhonors classes for the experiment, yielding 560 students in nonhonors classes and 265 in honors classes.

Table II presents tests of covariate balance. As expected given that randomization was across students within classrooms, the two groups are very well balanced on all measured dimensions, including sex, age, ethnicity, number of honors classes, and grade point average.¹⁵

III.B. Testing the Peer Pressure Mechanism

As noted already, any differences in the response to whether the sign-up decision is public or private across students in honors and nonhonors classes could arise for reasons other than differences in norms. For example, honors and nonhonors students may differ along many social, economic, and demographic dimensions or may have different aspirations or expectations, some of

15. We were able to get administrative data on honors classes and GPA for 94 percent of our sample. Missing information does not significantly correlate with treatment. Also, accessing individual GPA data requires both child assent and parental consent, which we did not receive from 16 percent of students. Therefore, we can only provide GPA data at the group level and cannot use it in our regressions. However, we also asked students to self-report grades on the survey handed out after the sign-up form was collected.

TABLE II
BALANCE OF COVARIATES

	(1) Private condition	(2) Public condition	(3) <i>p</i> -value
Male	0.506 [0.501]	0.518 [0.500]	.704
Age	16.74 [0.535]	16.75 [0.489]	.851
Hispanic	0.96 [0.196]	0.959 [0.2]	.899
# of honors classes	1.351 [1.486]	1.367 [1.477]	.880
GPA	2.52 [0.894]	2.48 [0.856]	.546
Observations	411	414	

Notes. Columns (1) and (2) report the mean level of each variable, with standard deviations in brackets, for the private and public conditions. Column (3) reports the *p*-value for the test that the means are equal in the two conditions.

which could affect how they respond to differences in whether information is private.

To reduce this heterogeneity and create a more comparable set of students in honors and nonhonors classes, which will allow us to estimate more cleanly the effect of changing just the composition of peers to whom the sign-up decision is potentially revealed, we can exploit the fact that many students do not take exclusively honors or nonhonors classes. In our sample schools, students are free to choose the honors or nonhonors version of every subject for which both are offered. Per school policy, they cannot be denied entry into any honors class that they want to take (even if they have poor grades), nor can they be forced to take an honors class they do not want to take. Many students therefore choose to take just a few honors classes, for example, choosing a subject that they are particularly interested in or a class with a teacher they like or heard good things about.

We can therefore examine students taking just some but not all honors classes and exploit variation in the timing of those courses relative to the timing of when our research team arrived to conduct the experiment. For any given some-honors student, whether the period when we arrived and conducted our study corresponded to one of their honors classes or one of their nonhonors classes should be exogenous with respect to their attributes. The effects of making sign-up public or private in honors

versus nonhonors classes for this group of students therefore more cleanly isolates how sign-up varies when essentially at random we offer it to them when they are sitting in the room with other honors students or other nonhonors students.

In practice, this strategy involves restricting our analysis to students taking between one and three honors classes.¹⁶ We note that in implementing this strategy, we must condition on the number of honors classes being taken so that we only compare students taking the same number. To see this, note that the full set of some-honors students we find in nonhonors classes will include a greater share of students taking just one honors class relative to students taking three honors classes (setting aside differences in the size of these two groups), since the former are much more likely to be in a nonhonors class during any given period. By contrast, the set of some-honors students we find in honors classes will contain a larger share of three-honors students relative to one-honors students. Since one- and three-honors students likely differ from each other in many ways, our empirical strategy relies on comparing only among those taking the same number of honors classes, who should therefore be similar, just exploiting variation in whether we happened to arrive when they were sitting in one of their honors classes or one of their nonhonors classes.

One potential concern is class scheduling. For example, suppose in the extreme case we visited only one school and that honors classes for the various subjects are offered uniquely across periods, that is, period 1 offers honors only in English, and honors English is only offered period 1. In this setting, if we arrived first period, the some-honors students found in an honors class will all be taking honors English, while those found in a nonhonors class will be taking honors only in other subjects. If students taking honors in different subjects differ from each other, particularly in ways that affect how they respond to whether their decisions are public or private (independently of peer pressure), then we will not rule out selection. Though we have no strong priors that such students would respond differently, we believe that in practice this is not a concern for our analysis. First, because these are large schools, there are multiple honors and nonhonors sections for each subject, offered

16. In our sample, taking four honors classes is effectively taking all honors (only nine students take five honors classes). Consistent with this, we find no four-honors students in any of the nonhonors classes we visited.

during different periods throughout the day. So visiting during one particular period will not necessarily skew the some-honors students we find in an honors class toward a particular honors subject relative to some-honors students we find in a nonhonors class. In addition, we visited each school during two separate periods. Finally, we visited different schools, each of which has different schedules (and we visited different schools during different periods).

This strategy assumes that information is to an extent localized. But a some-honors student sitting in an honors class when offered the course under the public regime may worry that peers in their nonhonors classes will learn that they signed up (especially since other students in the class are also taking a mix of honors and nonhonors classes). This would work against finding differences based on whether these students are offered the course when with their honors or nonhonors peers, and suggests we may underestimate peer pressure.¹⁷

Overall, there are 343 some-honors students (42 percent of our sample). Online Appendix Table A.V provides means for key covariates. We note that in columns (3) and (4) of Panel A, covariates are not balanced between those we surveyed in honors and nonhonors classes (though they are balanced across the public and private treatments). As discussed already, those we find in an honors class are taking more honors classes (and have a higher average GPA) than those we find in a nonhonors class. Thus, as a demonstration, Panel B shows means for the set of students taking exactly two honors classes (we focus on two-honors students because in practice with our smaller sample sizes we find very few one-honors students in an honors class and very few three-honors students in a nonhonors class, leading to small cell sizes and noisier estimates).¹⁸ Overall, the two groups are now very similar in terms

17. We cannot assess whether information flows across classes (or whether students believe it does). It is possible that students don't talk much about these kinds of efforts, and it is only when it is directly observed that it is relevant. There may also be a practice among some-honors students that "what happens in honors class, stays in honors class." For example, some-honors students may want to work hard and succeed in their honors classes, and may then worry that if they tell their nonhonors peers what another some-honors student did in an honors class, other students could in turn do the same to them.

18. We find only 13 one-honors students in an honors class, both because they don't take many honors classes and because we visited fewer honors classes, and only 9 three-honors students in a nonhonors class, both because they don't take many nonhonors classes and because there are fewer three-honors students.

of attributes (and covariates are again balanced across public and private treatments).¹⁹ They are also well balanced across honors subjects; of the two total honors classes they are taking, the groups differ only by 0.08, 0.02, and 0.12 in terms of the number of math/sciences, social sciences, and humanities honors classes. Though none of the differences are statistically significant, to absorb any residual variation, in separate regressions that follow we also add controls for attributes and honors subjects taken (this does not change the estimates appreciably).

III.C. Empirical Strategy

We begin by regressing an indicator for whether individual i in classroom c chose to sign up for the prep course (*Signup*) on an indicator for whether they were offered the public or private treatment (*Public*), separately for honors and nonhonors classes (pooling and using interactions yields similar conclusions).²⁰

$Signup_{i,c} = \beta_0 + \beta_1 Public_{i,c} + \varepsilon_{i,c}$, for class type = {honors, non-honors},

where β_1 , the estimated effect of making the sign-up decision public, is the coefficient of interest (separately for honors and nonhonors classes). In additional specifications, we add other covariates (age and dummies for sex and Hispanic) as well as surveyor and classroom fixed effects; the latter further isolate the within-classroom variation in the public versus private condition across students. These results will capture the overall effects of making sign-up public rather than private in the two types of classes, which can carry implications for school policies and practices.

To then more cleanly test the isolated peer pressure mechanism, we estimate the same regressions while limiting the sample to some-honors students, adding dummies for the number of honors classes to ensure that we only compare students taking the same number of such classes (and in additional specifications, we add controls for attributes and honors subjects).

19. One concern is that honors classes may be smaller than nonhonors classes, and peer pressure may differ by class size. However, in our sample, the difference is very small and not statistically significant.

20. We estimate separate regressions for boys and girls. There are small differences in behavior between boys and girls, but few are statistically significant.

III.D. *Main Results*

We begin by providing the raw sign-up rates across public and private conditions, in both honors and nonhonors classes. Figure IV displays the findings. In nonhonors classes, the private sign-up rate is 72 percent, while the public rate is 61 percent. The difference is significant at the 1 percent level. In honors classes, private and public sign-up rates are very high overall, and very similar: 92 percent of students sign up under the private treatment, while 93 percent sign up under the public one ($p = .631$). These high sign-up rates suggest that students indeed valued the course being offered, consistent with their beliefs about the cost of the course. The fact that sign-up is not affected by privacy in the honors class shows that there is no general effect of privacy itself (such as students always having a strong preference for greater privacy); though it is possible that the value placed on privacy differs between students in honors and nonhonors classes or that the demand for (or value of) the course is so much higher in honors classes (since more students want to go to college) that they are willing to accept the loss of privacy in exchange for the course. We separate out this possibility below.

In Table III, we present the results in regression format. Given the possibility of common shocks and correlated errors by classroom, along with the small number of classrooms in our sample, in addition to p -values from robust standard errors we also present p -values from wild bootstrap clustered standard errors (Cameron, Gelbach, and Miller 2008) and permutation tests.²¹ In column (1), we present the results without controls (which replicate the sign-up rates from Figure IV); in column (2) we add individual covariates and in column (3) we also add classroom and surveyor fixed effects. The results are very similar across specifications, suggesting that randomization was successful, and are robust to the various methods for computing standard errors. We again conclude that making sign-up public rather than private reduces sign-up rates in nonhonors classes, by a statistically significant 11–12 percentage points. But there is again no effect in honors classes. We believe these results are valuable in

21. For these tests, we randomly assign placebo treatment status (public sign-up) to different students and estimate placebo treatment effects. We repeat this process 1,000 times and compare the size of the true treatment effects to the distribution of placebo treatment effects.

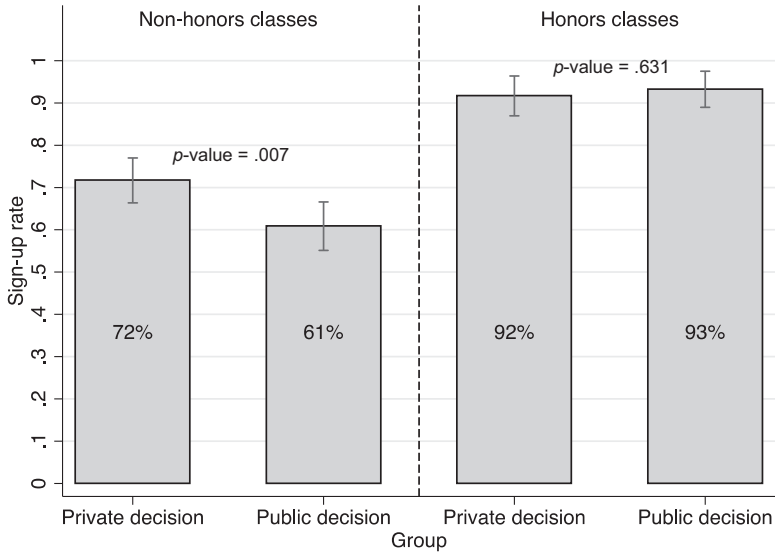


FIGURE IV
 Sign-up Rates for Private versus Public Decisions, NonHonors versus Honors Classes

This figure presents the means and 95 percent confidence intervals of the sign-up rates for students in the private and public conditions, separately for honors and nonhonors classes. There are 560 observations for nonhonors classes and 265 for honors classes.

themselves, aside from testing for peer pressure as the driving mechanism, with important implications for school policy and practices by showing a large, negative effect of observability on investment choices in nonhonors classes.

This first set of results indicates that there is not a universally negative effect of making the sign-up decision public. Nevertheless, this is not yet sufficient to establish the existence of different social norms in honors versus nonhonors classes, nor that students are responding to those differences. We therefore turn to our analysis of some-honors students. Having established that there are no significant differences between such students that were offered the SAT course in an honors or a nonhonors class (once we condition on the number of honors classes being taken), we can show that by contrast, their classmates in those classes are very different. In nonhonors classes, the private sign-up rate among no-honors classmates is 65 percent, whereas in

TABLE III
EFFECT OF PUBLIC TREATMENT ON SIGN-UP DECISION

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
	Dummy: student signed up for the SAT prep course					
Public treatment	-0.1083*** [0.040]	-0.1195*** [0.040]	-0.1231*** [0.040]	0.0157 [0.033]	0.0095 [0.032]	0.0092 [0.031]
Inference robustness (<i>p</i> -values)						
Robust standard errors	.007	.003	.002	.631	.766	.766
Wild bootstrap	.012	.006	.002	.596	.706	.720
Permutation test	.006	.001	.002	.480	.748	.749
Mean of private take-up		0.717			0.917	
Includes individual covariates	No	Yes	Yes	No	Yes	Yes
Includes classroom and surveyor fixed effects	No	No	Yes	No	No	Yes
Observations	560	531	531	265	258	258
<i>R</i> -squared	0.013	0.042	0.104	0.001	0.035	0.139
Sample	Nonhonors classes			Honors classes		

Notes. Columns (1) to (3) restrict the sample to nonhonors classes, and columns (4) to (6) restrict to honors classes. Columns (1) and (4) present OLS regressions of a dummy variable for whether the student signed up for the SAT prep course on a public sign-up dummy. Columns (2) and (5) replicate and add individual covariates (age and dummies for male and Hispanic). Columns (3) and (6) further add surveyor and classroom fixed effects. Robust standard errors in brackets. *** $p < .01$, ** $p < .05$, * $p < .1$. See Section III.D for a description of the wild bootstrap and permutation test procedures.

honors classes the rate among all-honors students is 100 percent (the p -value of the difference is .000). There are also dramatic differences in peers' GPA (2.03 in nonhonors versus 3.54 in honors, with $p = .000$). Some-honors students fall between the two, with 86 percent private sign-up rates and a 2.67 GPA.

These results establish that the peer groups are indeed very different in honors and nonhonors classes, and in a way that helps us formulate our hypotheses on the direction of social pressure effects for students taking some-honors classes. If peer pressure pushes students towards conforming to the locally prevailing norm within the classroom, we expect public sign-up to be lower than private sign-up in nonhonors classes, and higher in nonhonors classes.²² In Table IV, we estimate regressions using the full

22. Online Appendix Figure A.II shows the raw sign-up rates for two-honors students (again, for any presentation of means we must compare for a specific number of honors classes, and cell sizes for one- and three-honors students are small so the means are noisy). The figure provides striking preliminary visual

sample of some-honors students, separately for honors and non-honors classes. In nonhonors classes, the effect of the public treatment is to reduce sign-up rates by a statistically significant 15–17 percentage points. In honors classes, the public treatment increases sign-up rates by 7–9 percentage points, with statistical significance at the 10 percent level in three of the four specifications. Viewed in a different way, when choices are public, sign-up rates are over 20 percentage points greater when (otherwise identical) students make them in one of their honors classes rather than one of their nonhonors classes. The results are generally robust to the alternative forms of statistical inference, losing significance in just one case (column (5), where the p -value increases from .07 to .11 when using wild bootstrap clustered standard errors).

Of course, we cannot generalize the results for these some-honors students to all students (though the full sample results showing improved sign-up by making it private in nonhonors classes still holds). However, it is still valuable to document a set of students for whom the localized influence of peers can have such a dramatic effect. The set of some-honors students represents about 42 percent of the sample. Finally, these some-honors students may be the most relevant “marginal students”; those taking all honors classes are already making high levels of effort and investment, whereas those not taking any honors classes may require deeper interventions or altogether different policies to increase their effort.

III.E. Heterogeneity

Our main underlying hypothesis for why peer observability may affect choices is that students worry about what their peers will think of them. On a second form handed out to students after they had turned in the sign-up form, we asked students, “On a scale 1–5, how important do you think it is to be popular in your school? (1: not important...5: very important).” These are of course just subjective, self-reports, but they can provide suggestive corroborating evidence of our proposed mechanism. If the effects that we observe are driven by fear of social sanctions or seeking social approval, we would expect students who are more

evidence that the public treatment decreases sign-up in nonhonors classes dramatically, while increasing it in honors classes.

TABLE IV
EFFECT OF PUBLIC TREATMENT ON SIGN-UP DECISION FOR STUDENTS TAKING ONE TO THREE HONORS CLASSES

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Public treatment	-0.1673*** [0.061]	-0.1486** [0.061]	-0.1465** [0.063]	-0.1467** [0.064]	0.0850* [0.047]	0.0729 [0.046]	0.0834* [0.045]	0.0887* [0.048]
Inference robustness (<i>p</i> -values)								
Robust standard errors	.007	.017	.022	.023	.071	.115	.064	0.067
Wild bootstrap	.008	.008	.028	.032	.114	.106	.084	0.070
Permutation test	.006	.027	.030	.031	.061	.116	.069	0.046
Mean of private take-up			0.848			0.864		
Includes individual covariates	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Includes classroom and surveyor fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Includes honors subjects variables	No	No	No	Yes	No	No	No	Yes
Observations	184	176	176	176	159	155	155	155
Sample		Nonhonors classes			Honors classes			

Notes. This table restricts the sample to students taking one, two, or three honors classes. Columns (1) to (4) restrict the sample to nonhonors classes, and columns (5) to (8) restrict to honors classes. Columns (1) and (5) present OLS regressions of a dummy variable for whether the student signed up for the SAT prep course on a public sign-up dummy, controlling for dummies on the number of honors classes taken by the student. Columns (2) and (6) add individual covariates (age and dummies for male and Hispanic). Columns (3) and (7) further add surveyor and classroom fixed effects. Columns (4) and (8) replicate columns (3) and (7) adding controls for the number of honors classes taken by subject categories (math/sciences and social sciences; the omitted category is humanities). Robust standard errors in brackets. *** $p < .01$, ** $p < .05$, * $p < .1$. See Section III.D for a description of the wild bootstrap and permutation test procedures.

concerned with popularity to be more responsive to whether sign-up is public or private. To assess this hypothesis, we split our sample as close as possible to half, according to the importance attributed by students to being popular (answers 1 and 2 (not important) versus 3, 4 or 5 (important)). Figures V and VI present the results for the raw sign-up rates.²³ Figure V shows that for students in nonhonors classes who say that it is important to be popular, the sign-up rate is 20 percentage points lower in the public condition than in the private condition ($p = .002$). For those who care less about popularity, the effect of a public decision is small (4 percentage points) and not statistically significant. In Figure VI, we observe the opposite pattern for honors classes, although on a smaller scale (since the private take-up rates were already very close to 100 percent): a positive effect of public sign-up for those who care more about popularity, and no difference for those who care less. Table V presents the results in regression format, which confirm these results. Thus, we find that students who believe it is important to be popular move in the direction of locally prevailing norms (in both honors and nonhonors classes) when sign-up is public rather than private, while those who do not think it is important are unaffected by whether sign-up is public or private.²⁴

III.F. Account Login Data

Our main objective is to test for peer pressure, for which the sign-up decision is the relevant outcome. However, we also obtained data on whether students actually logged into the system later to activate their accounts (data on intensity of usage are not available). It is worth emphasizing that in analyzing this outcome, we lose experimental control because students in the public and private treatments are likely to have communicated or coordinated with each other after our team left the classroom. In doing so, they may have changed their beliefs about whether others would learn about their decision. Such communication also

23. The results and those that follow return to the full sample, since stratifying by popularity leads to extremely small cell sizes (popularity * honors/nonhonors * public/private) for the subset of some-honors students. The results for the some-honors students do show the same qualitative pattern but are less precisely estimated.

24. We can reject equality of the two effects under robust standard errors, but not when using the wild bootstrap clustered standard errors or the permutation test.

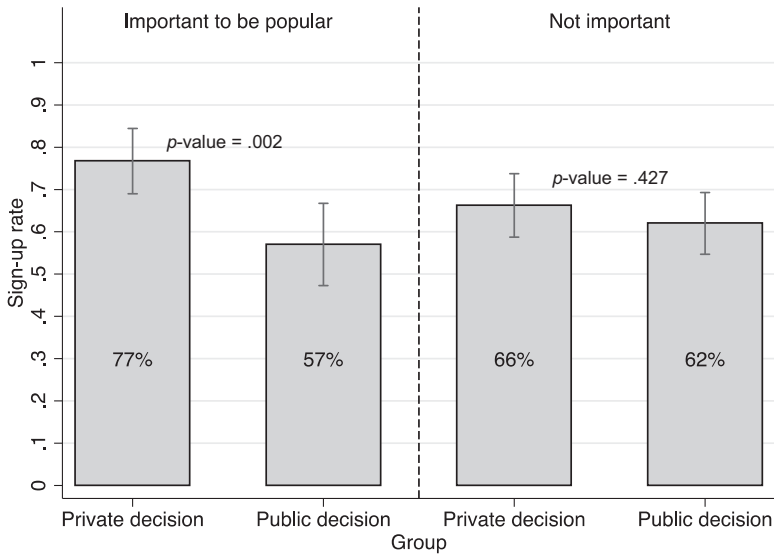


FIGURE V

Sign-up Rates for Private versus Public Decisions: Importance of Being Popular (NonHonors Classes)

This figure presents the means and 95 percent confidence intervals of the sign-up rates for students in the private and public conditions in nonhonors classes, separately for students who consider important to be popular in their school and those who do not. The dummy for whether the student considers it important to be popular is constructed by collapsing the answers to the question, “How important is it to be popular in your school?” from a 1–5 scale to a dummy variable (answers 3–5 were coded as considering it important, 1–2 as not important). There are 216 observations in the “important to be popular” panel and classes and 325 in the “not important” panel.

provides scope for other forms of peer effects beyond peer pressure, such as social learning or consumption externalities. So the estimates from this analysis are not as useful for testing our hypothesis. In addition, our analysis was designed to detect effects on sign-up rates, and we may therefore be underpowered to detect subsequent account login rates. However, activating the account is a useful policy outcome, indicating how much you can actually change adoption of an investment just by varying whether it is public or private. Examining this outcome can also help establish that signing up for the course was not just cheap talk, since we can see whether students at sign-up actually intended to follow through and use the course.

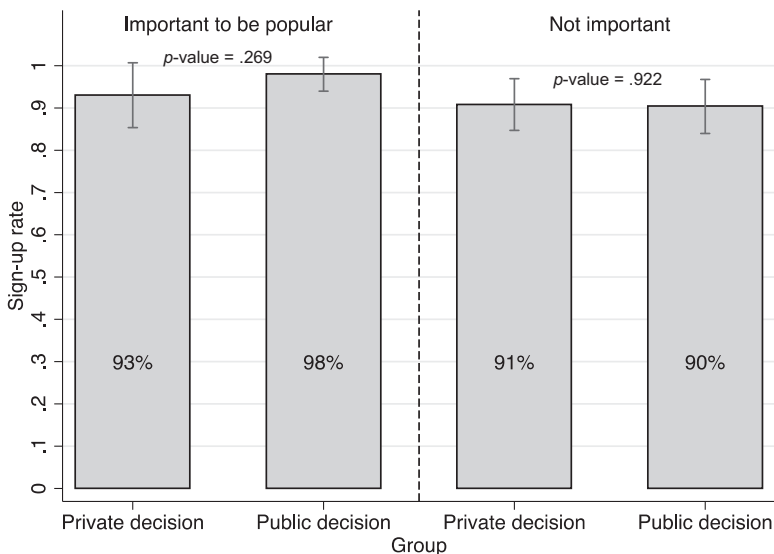


FIGURE VI

Sign-up Rates for Private versus Public Decisions: Importance of Being Popular (Honors Classes)

This figure presents the means and 95 percent confidence intervals of the sign-up rates for students in the private and public conditions in honors classes, separately for students who consider important to be popular in their school and those who do not. The dummy for whether the student considers it important to be popular is constructed by collapsing the answers to the question, “How important is it to be popular in your school?” from a 1–5 scale to a dummy variable (answers 3–5 were coded as considering it important, 1–2 as not important). There are 92 observations in the “important to be popular” panel and classes and 170 in the “not important” panel.

Overall, 81 percent of students who signed up for the course logged in to activate their account, which is a fairly high follow-through rate and again confirms that students valued the course. Overall, the unconditional mean take-up (login, conditional on being offered the course) is 61 percent. This is broadly similar to the 51 percent of students in our sample schools who take the SAT.²⁵ Students in honors classes had a slightly higher

25. The rate here is slightly higher, but there may be students in eleventh grade who still think they would like to go to college but who ultimately do not (because of performance, finances, or other factors).

TABLE V
EFFECT OF PUBLIC TREATMENT ON SIGN-UP DECISION: BY IMPORTANCE OF POPULARITY

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
	Dummy: student signed up for the SAT prep course					
Public*Not important to be popular (A)	-0.0425	-0.0518	-0.0483	-0.0044	-0.0220	-0.0215
	[0.053]	[0.054]	[0.054]	[0.045]	[0.043]	[0.043]
Public*Important to be popular (B)	-0.1972***	-0.2005***	-0.2156***	0.0494	0.0609	0.0605
	[0.063]	[0.064]	[0.062]	[0.044]	[0.047]	[0.047]
Important to be popular dummy	0.1049*	0.1347**	0.1480***	0.0222	0.0113	0.0084
	[0.055]	[0.055]	[0.055]	[0.050]	[0.053]	[0.051]
Inference robustness (<i>p</i> -values) for (A)						
Robust standard errors	.427	.335	.367	.922	.606	.619
Wild bootstrap	.440	.326	.364	.888	.582	.602
Permutation test	.448	.345	.385	.933	.612	.296
Inference robustness (<i>p</i> -values) for (B)						
Robust standard errors	.002	.002	.001	.264	.198	.197
Wild bootstrap	.066	.050	.024	.468	.364	.360
Permutation test	.002	.002	.000	.182	.175	.184
Mean of private sign-up for students who do not find it important to be popular		0.662			0.908	
Includes individual covariates	No	Yes	Yes	No	Yes	Yes
Includes classroom and surveyor fixed effects	No	No	Yes	No	No	Yes
Observations	541	521	521	262	256	256
<i>R</i> -squared	0.020	0.053	0.118	0.011	0.051	0.152
Sample	Nonhonors classes			Honors classes		

Notes. Columns (1) to (3) restrict the sample to nonhonors classes, and columns (4) to (6) restrict to honors classes. The dummy for whether the student considers it important to be popular is constructed by collapsing the answers to the question, "How important is it to be popular in your school?" from a 1–5 scale to a dummy variable (answers 3–5 were coded as considering it important, 1–2 as not important). Columns (1) and (4) present OLS regressions of a dummy variable for whether the student signed up for the SAT prep course on a public sign-up dummy, a dummy on whether the student considers it important to be popular in his or her school, and the interaction of the two. Columns (2) and (5) add individual covariates (age and dummies for male and Hispanic). Columns (3) and (6) further add surveyor and classroom fixed effects. Robust standard errors in brackets. *** $p < .01$, ** $p < .05$, * $p < .1$. See Section III.D for a description of the wild bootstrap and permutation test procedures.

follow-through rate (78 percent versus 84 percent), though the difference is not statistically significant.

The results are shown in Online Appendix Figure A.III (the conclusions from regressions are similar). For the full sample of students, we find that in nonhonors classes, making the course public reduces the rate of logging in to use the system by 8.2 percentage points (from a base of 57 percent when sign-up is private; p -value = .051). In honors classes, as with the sign-up decision, there is no difference in login rates between public and private

treatments (77 percent for private, 78 percent for public).²⁶ We also note that the follow-through rates did not differ across any of the (honors/nonhonors) \times (public/private) groups.²⁷

III.G. Other Concerns and Interpretations

One concern for external validity is that students may not have valued the course greatly (e.g., they believed the course was not very good, or they were already taking another course), and perhaps in settings with higher stakes, students are less affected by peer pressure.²⁸ However, we note that sign-up is extremely high when privacy from classmates was ensured. As noted already, follow-through rates for activation were very high. Finally, students estimated the cost at \$260, which is a high cost for these low-income students. Though of course it remains possible that for many students, the true value of this course was low.

With some investments that students may make in school, there is also the possibility that undertaking such efforts reveals low ability, such as the need for extra help or assistance. Of course, this is just one possible form of peer social concerns or pressure, or a micro-foundation for such behavior, and thus does not challenge our results. However, we believe that such effects are unlikely to underlie our results. SAT preparation, whether through books or classes, is very common, and not often associated with representing low ability. In our survey, students

26. For two-honors students (Panel B), the results are similar to those for sign-up, but less precisely estimated.

27. The follow-through rate for some-honors students in honors classes is 81 percent under the public treatment and 82 percent under the private treatment. Thus, the positive peer pressure effect observed above (increases in sign-up rates under the public treatment for some-honors students in honors classes) is unlikely to be just cheap talk, since they are just as likely to follow through and actually login and activate their account.

28. A related possibility is that students may have thought that they would have another chance to sign up later. We believe this is unlikely to account for our results. First, even if students believed they would have another chance, they would have to also believe that the later opportunity would differ on privacy. Second, since we concluded the study, no students who had not signed up communicated to our team (students took away forms with our contact information) or their teachers that they were interested in the course. Finally, we asked students from the last school we visited (after sign-up was complete) whether they believed they would have another chance to sign up, and 85 percent said no. This may even overstate the extent of such beliefs, since the act of asking the question may suggest or elicit that belief.

reported that they believed that on average about 43 percent (64 percent in honors classes) of their classmates were taking some other course to prepare for the SAT. In addition, honors students in our sample had very high sign-up rates (over 90 percent), suggesting that this is not a course only for the worst students.

Alternatively, students may not want to undertake efforts if final outcomes are also observable, such as due to a “fear of failure”: students who believe they have a high likelihood of failure on some observable outcome (such as getting into a good college or any college at all), may choose not to undertake effort (or even actively signal that they are not putting in effort) so that if they fail, others will believe it was because they did not try, rather than that they tried and failed. Again, we believe the asymmetric response to the public treatment makes this alternative less likely, since we would then need the effects to go different ways in different classes (i.e., some-honors students have a fear of failure in their nonhonors classes, but the reverse of the fear of failure when in their honors classes).²⁹

A final issue to consider is whether the effects are due to consumption externalities. Having more peers take the course (as might be expected in honors classes) may make the course more valuable because students can study together or learn from each other. The reverse would hold in nonhonors classes, where fewer peers are likely to take it. Though we cannot completely rule out this possibility, we believe it is unlikely to drive our results. Consider the some-honors students. If they believed that students in all classes would also be offered the prep course, then the full set of their friends who will be offered and take up the course, and thus the expected consumption externalities, should not differ based on whether they are sitting with their honors or nonhonors peers when they are offered the course. If these students instead believed that the course was only being offered to those in the class with them at that time, then under the private condition we should expect higher sign-up rates for those sitting in an honors class than for those sitting in a nonhonors class (since they should expect more of their honors class

29. Fear of failure could differ across settings. For example, students may fear failure more around nonhonors peers, who might mock them for even trying. On the other hand, fewer of their nonhonors peers will be going to good colleges or to college at all, so failing is not as stark a contrast as it might be compared to their honors peers.

peers to take it). However, as noted already, these private sign-up rates do not differ significantly. Thus, though there may be consumption externalities, students do not appear to act as though there are when they make their private sign-up decisions. In addition, we note that though consumption externalities on their own could explain a difference in sign-up rates in honors and nonhonors classes, it is less clear that it should affect differential sign-up within each class based on whether sign-up is public or private. However, we cannot rule out that beliefs about consumption externalities could differ within each class based on whether a student was in the public or private sign-up regime. This could arise if students themselves share our hypothesis; in other words, students given the public sign-up sheet in an honors class believe more of their classmates will sign up than students given the private sign-up sheet (and the reverse in nonhonors classes).³⁰

IV. CONCLUSION

We find evidence that student effort and investments are highly responsive to concerns about peer observability. We also find evidence suggesting that the results are specifically driven by concerns over the possibility of facing social sanctions or gaining social approval depending on effort or investments, or at least a desire to conform to prevailing social norms among peers in the classroom. The results have potential implications for understanding the nature and impact of peer interactions in the classroom more generally.

Though we are unable to link these changes in behavior to longer run education or labor market outcomes, the fact that we find similar results in two different settings suggests that such effects may be more widespread. In identifying this potentially important mechanism, we hope future work might explore whether similar effects are found where the long-run costs to

30. This will also depend on beliefs about whether the course was offered to all classes. A some-honors student in an honors class who gets the public sign-up may believe that more of their peers will sign up; but they may also think that same condition will reduce the number of peers that will sign up in their nonhonors classes (though they may be more likely to study with friends in their honors classes). So beliefs about the difference in the number of friends that will take the course may be ambiguous.

students are greater, as well as whether any policy interventions can help mitigate these effects.

ANDERSON SCHOOL OF MANAGEMENT, UCLA, AND NBER
WHARTON SCHOOL, UNIVERSITY OF PENNSYLVANIA, AND NBER

SUPPLEMENTARY MATERIAL

An Online Appendix for this article can be found at QJE online (qje.oxfordjournal.org).

REFERENCES

- Akerlof, George A., "Social Distance and Social Decisions," *Econometrica*, 65 (1997), 1005–1027.
- Ashraf, Nava, Oriana Bandiera, and Scott Lee, "Awards Unbundled: Evidence from a Natural Field Experiment," *Journal of Economic Behavior and Organization*, 100 (2014), 44–63.
- Austen-Smith, David, and Roland G. Fryer, Jr., "An Economic Analysis of 'Acting White,'" *Quarterly Journal of Economics*, 120 (2005), 551–583.
- Ausubel, David, *Maori Youth: A Psychoethnological Study of Cultural Deprivation* (North Quincy, MA: Christopher Publishing House, 1977).
- Bandiera, Oriana, Iwan Barankay, and Imran Rasul, "Social Incentives in the Workplace," *Review of Economic Studies*, 77 (2010), 417–458.
- Barankay, Iwan, "Rank Incentives: Evidence from a Randomized Workplace Experiment," Mimeo, University of Pennsylvania, 2012.
- Becker, Gary, and Kevin M. Murphy, "Social Forces, Preferences and Complementarity," in *Social Economics: Market Behavior in a Social Environment* (Cambridge, MA: Harvard University Press, 2000), 8–21.
- Berman, Eli, "Sect, Subsidy, and Sacrifice: An Economist's View of Ultra-Orthodox Jews," *Quarterly Journal of Economics*, 115 (2000), 905–953.
- Brown, B. Bradford, "The Extent and Effects of Peer Pressure among High School Students: A Retrospective Analysis," *Journal of Youth and Adolescence*, 11 (1982), 121–133.
- , "Adolescents' Relationships with Peers," in *Handbook of Adolescent Psychology* Richard M. Lerner and Laurence Steinberg, eds. (New York: Wiley, 2004), 363–394.
- Brown, B. Bradford, Donna Rae Clasen, and Sue Ann Eicher, "Perceptions of Peer Pressure, Peer Conformity Dispositions, and Self-Reported Behavior among Adolescents," *Developmental Psychology*, 22 (1986), 521–530.
- Bursztyjn, Leonardo, Florian Ederer, Bruno Ferman, and Noam Yuchtman, "Understanding Mechanisms Underlying Peer Effects," *Econometrica*, 82 (2014), 1273–1301.
- Cai, Jing, Alain De Janvry, and Elisabeth Sadoulet, "Social Networks and the Decision to Insure," University of Michigan Working Paper, 2012.
- Cameron, A. Colin, Johan B. Gelbach, and Douglas L. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 90 (2008), 414–427.
- Carrell, Scott E., Richard L. Fullerton, and James E. West, "Does Your Cohort Matter? Measuring Peer Effects in College Achievement," *Journal of Labor Economics*, 27 (2009), 439–464.
- Carrell, Scott E., Bruce I. Sacerdote, and James E. West, "From Natural Variation to Optimal Policy? The Lucas Critique Meets Peer Effects," *Econometrica*, 81 (2013), 855–882.
- Coleman, James, *The Adolescent Society: The Social Life of the Teenager and Its Impact on Education* (Glencoe, IL: Free Press, 1961).

- Conley, Timothy G., and Christopher R. Taber, "Inference with 'Difference in Differences' with a Small Number of Policy Changes," *Review of Economics and Statistics*, 93 (2011), 113–125.
- Dobbie, Will, and Roland G. Fryer, Jr., "Are High-Quality Schools Enough to Increase Achievement among the Poor? Evidence from the Harlem Children's Zone," *American Economic Journal: Applied Economics*, 3 (2011), 158–187.
- Duflo, Esther, Pascaline Dupas, and Michael Kremer, "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya," *American Economic Review*, 101 (2011), 1739–1774.
- Durlauf, Steven N., and Yannis M. Ioannides, "Social Interactions," *Annual Review of Economics*, 2 (2010), 451–478.
- Epple, Dennis, and Richard Romano, "Peer Effects in Education: A Survey of the Theory and Evidence," in *Handbook of Social Economics*, Vol. 1B, Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, eds. (Elsevier/North-Holland: Amsterdam, 2011), 1053–1163.
- Fordham, Signithia, and John U. Ogbu, "Black Students' School Success: Coping with the 'Burden of Acting White,'" *Urban Review*, 18 (1986), 176–206.
- Fryer, Roland G., Jr., "A Model of Social Interactions and Endogenous Poverty Traps," *Rationality and Society*, 19 (2007), 335–366.
- , "The Importance of Segregation, Discrimination, Peer Dynamics, and Identity in Explaining Trends in the Racial Achievement Gap," in *Handbook of Social Economics*, Vol. 1B, Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, eds. (Elsevier/North-Holland: Amsterdam, 2011), 1166–1191.
- Fryer, Roland G., Jr., and Lawrence F. Katz, "Achieving Escape Velocity: Neighborhood and School Interventions to Reduce Persistent Inequality," *American Economic Review*, 103 (2013), 232–237.
- Fryer, Roland G., Jr., and Paul Torelli, "An Empirical Analysis of 'Acting White,'" *Journal of Public Economics*, 94 (2010), 380–396.
- Gans, Herbert, *The Urban Villagers: Group and Class in the Life of Italian-Americans* (New York: Free Press, 1962).
- Jacob, Brian A., "Public Housing, Housing Vouchers, and Student Achievement: Evidence from Public Housing Demolitions in Chicago," *American Economic Review*, 94 (2004), 233–258.
- Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, 75 (2007), 83–119.
- Lee, Kenneth W., and William G. Warren, "Alternative Education: Lessons from Gypsy Thought and Practice," *British Journal of Educational Studies*, 39 (1991), 311–324.
- Mas, Alexandre, and Enrico Moretti, "Peers at Work," *American Economic Review*, 99 (2009), 112–145.
- Oreopoulos, Philip, "The Long-Run Consequences of Living in a Poor Neighborhood," *Quarterly Journal of Economics*, 118 (2003), 1533–1575.
- Postlewaite, Andrew, "Social Norms and Social Assets," *Annual Review of Economics*, 3 (2011), 239–259.
- Sacerdote, Bruce, "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *Quarterly Journal of Economics*, 116 (2001), 681–704.
- Santor, Darcy A., Deanna Messervey, and Vivek Kusumakar, "Measuring Peer Pressure, Popularity, and Conformity in Adolescent Boys and Girls: Predicting School Performance, Sexual Attitudes, and Substance Abuse," *Journal of Youth and Adolescence*, 29 (2000), 163–182.
- Tran, Ahn, and Richard Zeckhauser, "Rank as an Inherent Incentive: Evidence from a Field Experiment," *Journal of Public Economics*, 96 (2012), 645–650.
- Zimmerman, David J., "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *Review of Economics and Statistics*, 85 (2003), 9–23.

This page intentionally left blank