The Network Origins of Pluralistic Ignorance: Theory and Evidence from Study Effort Stigma

Vasu Chaudhary, Krishna Dasaratha, William Murdock III, and Dev Patel

September 29, 2026*

Abstract

What drives misperceptions about what others think? We study this question in the U.S. school system, where peer pressure about academic achievement shapes study effort. Students misperceive this stigma: 94% overestimate the likelihood their peers dislike socializing with academically-inclined students. To explain this pattern, we build a network formation and social learning model that generates pluralistic ignorance via the *friendship paradox*: on average, someone's friends have more friends than they do. Consistent with the model's predictions, we find that those with more connected friends overestimate stigma the most and that students systematically underestimate the true study effort of their more popular friends. A randomized intervention correcting the resulting pluralistic ignorance decreases students' stated desire to hide their academic performance by 36%.

*Chaudhary (chaudhary.vasu.27@gmail.com, Northwestern University); Dasaratha (krishnadasaratha@gmail.com, Boston University); Murdock (william.murdock@gmail.com, Lazard); Patel (dev_patel1@brown.edu, Brown University). We are grateful to Alex Hayes for helpful feedback. This project was supported by Character Lab and facilitated through the Character Lab Research Network, a consortium of schools across the country working collaboratively with scientists to advance scientific insights that help kids thrive. This study was pre-registered here. This research underwent ethics review by the Committee on the Use of Human Subjects at Harvard University.

1 Introduction

Beliefs about what others think play an important role across a host of economic domains, from investing in stocks (Allen et al., 2006) to social learning about important policies (Banerjee et al., 2024). When those beliefs are incorrect, mistaken perceptions about the views of those around us can distort behavior in important ways (Bursztyn and Yang, 2022), such as influencing intra-household constraints about women's labor force participation (Bursztyn et al., 2020a) or the choice to engage in political protest (Cantoni et al., 2019).

What drives misperceptions about what others think? We examine the scope for people to learn from those around them yet nevertheless fail to converge to the truth. In our framework, the process of social learning not only fails to correct pluralistic ignorance but in fact generates bias in the first place.

We begin by building a model of endogenous network formation and learning. People form beliefs by observing their neighbors' behavior. When the characteristic of interest is correlated with the propensity to form links, then neighbors' types exhibit bias via the the *friendship paradox* (Jackson, 2019): on average, someone's friends have more friends than they do.

We test this model empirically in the context of U.S. middle schools, where peer pressure can shape study effort (Fryer and Torelli, 2010; Bursztyn and Jensen, 2015; Bursztyn et al., 2019). Norms about the social status of high academic achievement capture all of the classical features of the economic relevance of second-order beliefs: students make important decisions (how much to study) based in part on their perceptions about how *others* think about that action. If students hold incorrect beliefs about this social penalty, they may behave sub-optimally.

In this context, the forces of the model operate through the fact that students perceive popularity and study effort as substitutes: they believe that people who study more have less time for socializing. As a result, when students form beliefs about relevant social norms in the school from their direct friends, they place disproportionate weight on peers who socialize more and are perceived to study less. This generates systematic overestimation of stigma around study effort.

Empirically, we document this substantial pluralistic ignorance about studying norms. To establish this pattern, we survey 722 middle school students from a large, diverse, suburban and urban public school district in Florida in January 2022. We

measure peer pressure about studying using students' responses to a question about a "typical student" at their school. Specifically, we ask, "If you found out this student studied two hours more per week than you thought, how would that impact how much you would want to be friends with them? Everything else you know about the student is still the same. For example, they still have the same amount of time to spend with friends and family or to play sports." The rationale behind this phrasing is to encourage students to consider a shock to a peer's study habits, holding all else constant about that student. In particular, we do not want students to make other inferences about the student, such as their availability for socialization, but rather isolate preferences for friendship with respect to study effort. We define the social norm as the share of students answering "much less" or "slightly less likely want to be their friend," which we refer to as "negative peer pressure."

We find evidence that students significantly overestimate the stigma towards study effort held by their peers. Overall, 6.9% of students in our data answered that they would be less likely to be friends with someone after learning they put in more study effort. The vast majority of students overestimated this social pressure against studying. Across all students, 94% overestimate the share of their peers who report being less likely to be friends with another student who studies more. These errors are large in magnitude: the average overestimation is 39.25 percentage points. The size of these errors are unlikely to be explained by other factors such as experimenter demand or features of the survey elicitation: instead, we interpret these patterns as strong evidence that the majority of students privately hold relatively favorable views toward academic effort but incorrectly believe that others do not.

Beyond this headline pluralistic ignorance result, the model generates several other predictions and necessary assumptions that we test in our data. For several of these, we leverage social network data that we collected as part of the survey. Due to privacy concerns and logistical constraints in administrating the instrument, a full network mapping was not possible: instead, we use aggregated relational data (ARD) to estimate social network statistics for each student (Breza et al., 2020; Reeves et al., 2024; Breza et al., 2023).

We examine how student's beliefs about peer pressure in their school vary with the network position of their peers. Consistent with the model's predictions, students with higher degree neighbors perceive significantly greater academic stigma in their school environment. These students also underestimate the study effort of their more socially

connected peers, with the magnitude of this underestimation increasing in neighbors' degree. Interestingly, these belief patterns do not reflect underlying differences in true behavior: contrary to students' perceptions about how "cool" students spend their time, we do not find that high degree students exert significantly less effort, as measured by reported study hours or grades.

Finally, we conclude by verifying the consequences of pluralistic ignorance through a simple, light-touch information intervention. We provide students with ground-truth estimates of the true attitudes of their peers based on the other survey responses that we collected. Specifically, treated students were shown what percentage of students who were taking the same survey reported that they would be less likely to be friends with someone who studies more. Students assigned to control were instead asked to answer a free-response question about whether or not they believe studying impacts a student's popularity. We find that receiving information lowers students stated willingness-to-pay to hide their grades from their peers by 50 percent ($p-$value $< .01$), an effect that is larger for students with more mistaken beliefs initially.

Together, our results shed light on the origins of mistaken beliefs about the views of others. We bring empirical evidence to bear on a simple model of social learning to explain pluralistic ignorance in an economically important setting.

Related Literature Our paper builds on a rich literature examining the origins and consequences of social norms.[1] In school settings, a series of papers have documented how one's peers influence human capital decisions via a social mechanism. Bursztyn and Jensen (2015) show that educational investments strongly react to the prevailing norm in U.S. high schools, whether it is in support of studying more or less (e.g. in honors vs. non-honors classes). Bursztyn et al. (2019) illustrate how peer pressure can discourage visible investments in very different settings through a desire to hide either effort in one or low ability in another. Fryer and Torelli (2010) explore the racial dimension of peer pressure by conducting an empirical analysis of "acting white" in which educational investments signal one's opportunity costs of peer-group loyalty. In a field experiment in Chicago high schools, Ye (2022) finds that high-achieving students are particularly influential in shaping social norms about participation. Our paper examines the accuracy and origins of these student perceptions of peer pressure

[1] See Tankard and Paluck (2016), Bursztyn and Jensen (2017), and Bursztyn and Yang (2022) for excellent reviews on these and closely related topics.

in a diverse educational setting.

The potential for inaccurate beliefs like these to shape important economic decisions has been shown across a wide variety of settings, including but not limited to support for public policies (Alesina et al., 2022; Haaland and Roth, Forthcoming), political protests (Cantoni et al., 2019), school enrollment (Arteaga et al., 2022; Jensen, 2010), college majors (Zafar and Wiswall, 2015; Hastings et al., 2015; Conlon, 2020; Conlon and Patel, 2022), and anti-American attitudes (Delavande and Zafar, 2018). Like the perceptions of peer pressure we document in this paper, a growing literature has documented other instances of "pluralistic ignorance" in which one's in actuality majority beliefs is wrongly perceived to be the minority view, leading to incorrect views about the average belief in the community in settings such as women's labor force participation in Saudi Arabia (Bursztyn et al., 2020a) and India (Bernhadt et al., 2018), xenophobia in the United States (Bursztyn et al., 2020b), and political correctness on college campuses (Braghieri, 2021). In this paper, we combine individual-level beliefs with rich social network data to investigate the origins of these mistaken perceptions and shed light on the role of key, central figures in the social network in shaping the community-level norm.

The critical role of social networks in shaping economic behavior has been established across many settings, including education (Calvó-Armengol et al., 2009; Bramoullé et al., 2020). By linking these networks to beliefs about oneself and others (Jackson, 2019; DeGroot, 1974; Friedkin and Johnsen, 1990; DeMarzo et al., 2003), we illustrate an important channel through which social structures can influence community norms.

2 A Model of Belief Formation with Endogenous Networks

We next describe our model of belief formation. Suppose agents form networks endogenously and then form beliefs about the distribution of a characteristic based on observing neighbors' behavior. If the characteristic of interest is correlated with propensity to form links, then neighbors' types will be biased due to the *friendship paradox*: on average, agents' friends have more friends than they do. If agents treat their friends as a representative sample, this will lead to biased average beliefs about

the relevant characteristic. We begin with a model capturing this intuition in a single period of updating (as in Jackson 2019) and then extend the forces to a setting with many rounds of social learning (based on the Friedkin and Johnsen 1990 belief updating dynamic).

2.1 Misperceptions with One-Shot Updating

We begin with a model with a single round of updating. Agents allocate time between socializing and a task requiring effort. These choices generate a network according to a standard random network formation model, with degrees depending on time spent socializing. Agents then observe their neighbors' effort and form beliefs about peers' preferences. We now describe the details.

Each agent has a time budget $T > 0$ which they allocate between socializing and effort (which will correspond to time spent studying). We let the time devoted to effort be t_{i1} and the time devoted to socializing be t_{i2}. Each agent also has a type θ_i which captures how much the agent values effort. Each agent's type $\theta_i = \mu_s + \epsilon_i$ is drawn i.i.d. from a distribution with mean μ_s and mean-zero error ϵ_i. The mean $\mu_s \sim \mu + \epsilon_s$ in the network is drawn i.i.d. from a distribution with mean μ and mean-zero error ϵ_s.

Each agent i knows their own θ_i and chooses $t_{i1} \geq 0$ and $t_{i2} \geq 0$ to maximize a utility function $u_i(t_{i1}, t_{i2}, \theta_i)$ that is strictly increasing in the first two arguments and twice-differentiable, subject to $t_{i1} + t_{i2} \leq T$. We assume that

$$\frac{\partial^2 u_i}{\partial t_{i1} \partial \theta_i} > \frac{\partial^2 u_i}{\partial t_{i2} \partial \theta_i},$$

$u_i(t_{i1}, T - t_{i1}, \theta_i)$ is strictly concave in t_{i1} for each θ_i, and $u_i(0, T, \theta_i) = u_i(T, 0, \theta_i) = -\infty$. These assumptions imply that the optimal t^*_{i2} is unique and is strictly decreasing in θ_i.

We next describe the network-formation protocol. An agent who spends time t_{i2} socializing has putative network degree $\widetilde{d}_i \sim F_{t_{i2}}(d)$ drawn independently from a distribution depending on t_{i2}. We also assume that $F_t(0) > 0$ for all t.[2] We assume

[2]This will ensure that there is a positive probability of generating a graphical sequence, and could be replaced by other assumptions ensuring this.

that the family of distributions satisfies the strict monotone likelihood ratio property:

$$\frac{F_t(d)}{F_t(d')} > \frac{F_{t'}(d)}{F_{t'}(d')}$$

whenever $t > t'$ and $d > d'$. A natural example is for the putative degrees to be drawn from Poisson distributions with means t_{i2}.

We want to obtain an actual degree sequence that is *graphical*: there is a network with this degree sequence (see Erdös and Gallai 1960 for necessary and sufficient conditions). To do so, we simultaneously redraw all putative degrees until we obtain a graphical sequence of degrees $(\widetilde{d}_1, ..., \widetilde{d}_n)$, which will be the degree sequence for the network between agents. We refer to these realized degrees as $(d_1, ..., d_n)$.

Given degrees $(d_1, ..., d_n)$, we match agents according to the configuration model: select an undirected network uniformly at random among all networks with n nodes and degree distribution $(d_1, ..., d_n)$. We denote the set of i's friends by N_i.

Each agent i knows the utility functions u_j and observes the socializing time t_{j2} of each friend $j \in N_i$. By our assumption that t^*_{j2} is strictly decreasing in θ_i, student i can infer each θ_j from the observed socializing times t_{j2}. Each agent estimates μ_s as the average

$$\hat{\mu}_{s,i} = \frac{1}{|N_i|} \sum_{j \in N_i} \theta_j.$$

In particular, the agent fails to adjust for the friendship paradox and treats the neighborhood N_i as a representative sample from the network. This leads to misperception about μ:

Proposition 1. *The expected belief about μ_s is smaller than μ:*

$$\mathbb{E}_i[\hat{\mu}_{s,i}] < \mu.$$

The proposition states agents underestimate how much their peers value effort. The intuition is simple: more connected agents disproportionately affect perceptions, and because studying and socializing are substitutes these students tend to value studying less. The proof uses a coupling argument to formalize this intuition in the configuration model. The result is similar to the analysis sketched out in Section IIIC of Jackson (2019), though we focus on beliefs rather than equilibrium behavior with complementarities.

In our empirical setting, the result will translate to students underestimating how much their peers value studying. We formalize this in a series of testable assumptions and predictions, which we will examine empirically in Section 3.

First, the network-formation model generates a negative correlation between value of studying and number of friends. The true relationship is not essential for the model predictions, but we do require that agents perceive such a negative correlation when forming beliefs:

Model Assumption 1. *Students believe that peers with a higher value of studying have fewer friends.*

Suppose students observe friends' time spent socializing and infer study hours based on this perceived correlation. Then Proposition 1 shows students underestimate average time spent studying:

Model Prediction 1. *If Assumption 1 holds but study hours are actually independent of number of friends, then on average students underestimate friends' study hours.*

Moreover, this misperception is driven by students whose neighbors have more friends:

Model Prediction 2. *If Assumption 1 holds, then students with higher average neighbor degree underestimate friends' study hours more.*

We can say more if students form beliefs about social norms around studying based on their beliefs about others' behavior, as captured by the following assumption:

Model Assumption 2. *Students believe that peers with a higher value of studying have lower stigma.*

When this assumption additionally holds, the predictions about studying hours both have analogs about stigma:

Model Prediction 3. *If Assumptions 1 and 2 hold, then average beliefs about stigma are higher than true stigma.*

Model Prediction 4. *If Assumptions 1 and 2 hold, then students with higher average neighbor degree believe stigma is higher.*

Unlike the first prediction, the predictions about stigma do not depend on the true relationship between study hours and degree.

We end this section with several remarks on the results. First, Bursztyn and Yang (2022) document misperceptions in a wide range of settings and find these misperceptions are often asymmetric. Our friendship paradox-based mechanism generates such asymmetry and makes an explicit prediction about its direction: beliefs are biased toward the behavior that is correlated with higher sociability.

Second, our model assumes a frequentist belief formation procedure, but the same force would arise in a misspecified Bayesian model where students treat their friends as a representative sample from the school. We briefly sketch such a model. Suppose that the distributions $\mu_s \sim \mathcal{N}(\mu, \sigma^2)$ and $\theta_i \sim \mathcal{N}(\mu_s, \sigma_s^2)$ are Gaussian. If students treat their friends as a representative sample but are otherwise Bayesian, they form beliefs

$$\hat{\mu}_{s,i} = (1 - (d_i + 1)w_i)\mu + w_i\theta_i + \sum_{j \in N_i} w_i\theta_j,$$

where $w_i = \frac{\sigma_s^{-2}}{\sigma^{-2} + (d_i+1)\sigma_s^{-2}}$. The third term captures the same main force as our model: the same argument as in the proof of the proposition shows it is biased downward. The second term complicates the story, however: the weight w_i is lower when θ_i is smaller, which biases $w_i\theta_i$ in the opposite direction.

2.2 Misperceptions with Social Learning

Do these misperceptions persist after more interaction? The belief-updating rule in the previous subsection suggests analyzing many periods of social learning following DeGroot updating ((DeGroot, 1974)). A salient feature of our data, however, is persistent dispersion in beliefs and actions rather than the consensus predicted by DeGroot updating. We therefore study a version of the naive learning rule of Friedkin and Johnsen (1990), which modifies DeGroot updating so that agents continue to place positive weight on their types.

The model consists of an initial network-formation period $t = 0$ and an infinite sequence of social-learning periods $t = 1, 2, \ldots$. In each period, each agent chooses an action $a_{i,t}$, which we interpret as time spent socializing. Agent types are drawn as in the one-shot model.

In period $t = 0$, agents choose actions $a_{i,0}$ to maximize $u_i(a_{i,0}, \theta_i)$. We assume

that the optimal choice of $a_{i,0}$ is unique and strictly decreasing in θ_i, and assumptions in the one-shot model above are sufficient to ensure this. Given actions, a random network is drawn from the configuration model as in the one-shot model (with time spent socializing now denoted by $a_{i,0}$ rather than t_{i2}).

In each period $t \geq 1$, agents form beliefs $\hat{\mu}_{i,t}$ about the average type in society as follows: agents infer the types $\theta_{j,t-1}$ for each neighbor that would generate their actions $a_{j,t-1}$ and set

$$\hat{\mu}_{i,t} = \frac{1}{|N_i|} \sum_{j \in N_i} \theta_{j,t-1}.$$

We can capture these weights in a matrix W with entries $W_{ij} = \frac{1}{|N_i|}$ if $j \in N_i$ and $W_{ij} = 0$ otherwise. In our structural estimation, we will allow a larger set of possible weights.

In periods $t \geq 1$, agents choose actions to maximize $u_i(a_{i,t}, \theta_{i,t})$, where

$$\theta_{i,t} = \lambda \theta_i + (1 - \lambda)\hat{\mu}_{i,t}.$$

Agents behave as if their type is a convex combination of their underlying type and their belief about the average type in the population, capturing a desire to conform to others' behavior. The parameter $\lambda \in (0, 1)$ captures the weight on own types.

We characterize long-run beliefs and actions and show that beliefs are biased as in the one-shot model:

Proposition 2. *Long-run beliefs are*

$$\hat{\mu}_\infty := \lim_{t \to \infty} \hat{\mu}_t = \lambda(I - (1 - \lambda)W)^{-1} W\theta.$$

Agents choose long-run actions as if their types are

$$\theta_\infty := \lim_{t \to \infty} \theta_t = \lambda\theta + (1 - \lambda)\hat{\mu}_\infty.$$

The long-run beliefs are therefore sums of terms proportional to $W^k\theta$ for $k \geq 1$. Proposition 1 showed the term $W\theta$ is biased downward because higher-degree agents receive higher weights, and the higher-order terms $W^k\theta$ for $k > 1$ are similarly biased. So the misperceptions about the type distribution in the population persist through repeated naive updating. This contrasts the naive learning results of Golub and Jackson (2010), which require initial opinions that are independent of network

position.

3 Testing the Model's Predictions

We now turn to U.S. middle schools to test the predictions of this framework empirically in the context of peer pressure about study behavior.

Data Collection We surveyed students from a large, diverse, suburban and urban public school district in Florida in January 2022.[3] After dropping students due to technical issues in the survey implementation and those who fail our attention checks, our final sample includes 772 students. About one quarter of of this set qualify for free or reduced price lunch, 37% are Hispanic, and 12% are Black. Appendix Table C.1 compares this final sample to all students assigned to take the survey. We also received administrative data from the school district that includes traditional demographic variables and transcript information.

Measuring Students' Social Networks To measure the social connections among students in our sample, we combine administrative data on the universe of students in each school with aggregated relational data (ARD) to estimate students' social networks (Breza et al., 2023; Reeves et al., 2024).[4] Concretely, survey questions take the form: *"How many students in your school do you know who have trait t?"* asked over a set of traits provided in Appendix C.2. We use these ARD responses, together with population trait prevalence from administrative records, to estimate a network formation model. Students report that 90% of their friends share the same grade, and thus we estimate a graph within each school-grade.

We consider three approaches to estimating the network and report results using all three methods. Appendix Section C.3 provides further technical details of the estimation approach. First, we consider a stochastic block model. Second, we apply a spectral estimation approach following Hayes et al. (2026). Finally, we estimate a geometric network formation model following Breza et al. (2020) using Markov chain

[3]For further details on the data collection process and our implementing partnership with the Character Lab Research Network (CLRN), see Appendix Section C.1. Appendix C.2 reproduces our survey instrument in its entirety.

[4]Restrictions imposed by the school district and our implementing partner prevented us from directly eliciting friendships and from collecting data from the full social network.

Monte Carlo estimation, denoted "MCMC" in the tables.

3.1 Testing Model Assumption 1

We first examine whether students believe that their peers who value studying less also have more friends. We ask respondents, "How many hours more/less do students think popular kids study compared to a typical student?" Figure 1 presents a histogram of answers to this question. The vertical dashed line marks 0—the point at which respondents expect popular students and typical students study the same amount. Most of the mass of the histogram falls to the left of this point, indicating that the majority of students think that popular students study less than their typical peers. The mean gap is about two hours a week.

3.2 Testing Model Prediction 1

We examine whether students underestimate the study hours of their peers. We compare students' perceived average friend study hours to the average study hours of their friends recovered from the simulated network draws. The evidence again supports the model's mechanism. On average, students underestimate their friends' study effort. The mean difference between actual and perceived average friend study hours is positive though not statistically significant at conventional levels (0.16 hours, p-val = 0.15), and 63 percent of students understate how much their friends study.

3.3 Testing Model Prediction 2

We test whether students with higher degree friends underestimate their peer study effort more. Table 2 presents results of this regression. In the baseline specification, a one-unit increase in average neighbor degree is associated with 0.31 hours per week reduction in perceived friend study effort. One would expect that belief about friends' study hours would be driven by one's own study effort as well. In column 4, even after we control for own study effort, increased average neighbor degree still remains negative and significant. In Table A.1 we look at the extensive margin and see that one unit increase in average neighbor degree is associated with 4.7 percentage point increase in the probability that a student underestimates their friends' study effort.

3.4 Testing Model Assumption 2

Next, we turn to whether students perceive that those with a lower value of studying have higher stigma about those who study more. Among a random half of respondents (those in the control group of the information experiment, described below), we ask a free response question: "Do you think studying impacts a student's popularity? Why or why not?" We classify views on this relationship using an LLM and by hand-checking a sample to verify those classifications. Figure 2 presents results of this exercise. Slightly under half of students say there is no link between study effort and popularity—the modal answer. The next most common response, consistent with the model's assumption, is a negative relationship, identified by about a quarter of respondents. Only 8.9% of students say they perceive a positive association.

3.5 Testing Model Prediction 3

Under the two assumptions above, students should exhibit pluralistic ignorance about social stigma regarding studying. To test for mistaken beliefs about others, we first establish a ground truth by measuring peer pressure about studying using students' responses to a question about a "typical student" at their school. Specifically, we ask, "If you found out this student studied two hours more per week than you thought, how would that impact how much you would want to be friends with them? Everything else you know about the student is still the same. For example, they still have the same amount of time to spend with friends and family or to play sports." The rationale behind this phrasing is to encourage students to consider a shock to a peer's study habits, holding all else constant about that student. In particular, we do not want students to make other inferences about the student, such as their availability for socialization, but rather isolate preferences for friendship with respect to study effort. We define the social norm as the share of students answering "much less" or "slightly less likely want to be their friend," which we refer to as "negative peer pressure." Overall, 6.9% of students in our data answered that they would be less likely to be friends with someone after learning they put in more study effort.

Consistent with Prediction 3, the vast majority of students overestimated this social pressure against studying. Figure 3 presents the distribution of student estimates about the share of other respondents who exhibit negative peer pressure. As students typed numbers into this question, an array of boxes became shaded to reflect their

answers as a way of visualizing their answer and to help ensure comprehension of this type of belief elicitation. The vertical red line in Figure 3 denotes the truth. Across all students, 94% overestimate the share of their peers who report being less likely to be friends with another student who studies more. These errors are large in magnitude: the average overestimation is 39.25 percentage points. These gaps appear across demographic groups, as shown in Appendix Table A.2. Most students appear to hold relatively favorable views toward academic effort, but incorrectly believe that others do not.

Robustness of Peer Pressure Misperception We probe several potential alternative explanations for these patterns. First, one may be concerned that the question asks about a "typical student" at the respondent's school. Due to the small number of observations within most schools based on the survey sampling, we are underpowered to reliably estimate school-specific averages and thus focus on the overall response. Nevertheless, the data suggest this does not drive the misperception result. The school each student attends does little to predicts the distribution of beliefs: school fixed effects explain just 1.1 % of the overall variation. Second, because we drop students who do not pass attention checks, we may have a selected sample of responses. Yet the magnitudes of the errors far exceed anything that could be explained by the composition of students who pass the sample restrictions. Third, students' reported beliefs are likely subject to non-classical measurement error such as round number bunching and probability weighting, but again these biases cannot explain the magnitudes. Fourth, the question uses the term "typical" to be most accessible to the students taking the survey. We have focused on average responses when translating "typical" into a statistical moment, but results remain nearly identical when using the median or mode instead.

Experimenter Demand Experimenter demand could certainly bias this estimate downwards: for instance, even though survey responses were completely anonymous, students may have felt pressure to report that they would be happy to be friends with anyone, even if they actually felt otherwise. This challenge plagues all survey measures of pluralistic ignorance, especially on stigmatized topics. For the purposes of this study, the exact magnitudes do not play a key role—rather, our story relies on the existence of a gap, even if smaller than what the raw numbers may imply, and

the correlation with network structure.

3.6 Testing Model Prediction 4

We next examine whether students with higher average neighbor degree perceive more stigma—a key prediction of the social network underpinnings of pluralistic ignorance.

Table 1 regresses students' prior belief about school-wide stigma on the average degree of their neighbors under each of the three different network estimation approaches. The coefficient on average neighbor degree is positive and statistically significant, and the relationship remains positive even when we control for student's own degree. A one-unit increase in average neighbor degree is associated with a 2.09 percentage point (4.79%) increase in perceived stigma. In contrast, the coefficient on own degree is much smaller in magnitude. This distinction is important for our mechanism. Our model does not require that a student's own popularity predict their beliefs about school-wide stigma. Rather, what matters is the connectedness of the peers from whom they infer the broader social environment.

4 The Consequences of Misperceptions

To examine the implications of these misperceptions, we conduct a simple, light-touch information experiment to correct students' pluralistic ignorance about study norms. As our main outcome, we focus on students' self-reported willingness-to-pay (WTP) to conceal their grades from their peers.[5]

Experimental Design We randomly assign students to treatment or control $Treatment \in \{0, 1\}$. Within the treatment group, we provide information on their peers' attitudes about other students' studying habits. Specifically, treated students were shown what percentage of students who were taking the same survey reported that they would be less likely to be friends with someone who studies more. Students assigned to control were instead asked to answer a free-response question about whether or not they believe studying impacts a student's popularity.

[5]Due to restrictions from the school district and implementing partner, we were not able to incentivize any outcome for this experiment.

To calculate the information provided to each treated student, we randomly drew 30 responses from the set of students who had already completed the survey. We assign the first 30 respondents to complete the survey to control. Appendix Figure A.3 shows the distribution of treatment values.

We estimate the treatment effect of information provision using the specification in equation 1, where T_i indicates whether student i received the information and $\mathbf{X}_i$ denotes a vector of controls including school fixed effects, four GPA bins, 10 bins of baseline beliefs, student's self-reported study hours, and gender.

$$Y_i = \alpha + \beta T_i + \mathbf{X}_i \Gamma + \varepsilon_i \tag{1}$$

Table 4 reports the results of these regressions, all of which include robust standard errors. Those shown information about the true amount of social stigma in the survey say they are less willing to pay to conceal their grades. The magnitude is large: off of a control mean of \$17.95, treatment reduces WTP by \$7.30. We do not see strong treatment effect heterogeneity by students' prior. Under our model, priors about social stigma are endogenous. The heterogeneity results are consistent with non-random selection of students by baseline beliefs that shapes how they respond to information.

One concern with these results is experimenter demand: students may simply be reporting what they think the researchers want to hear, and those in the treatment group may have differential views about this. While we cannot fully rule this channel out, we can show in placebo tests that the treatment did not impact other outcomes that may be socially desirable along the same dimension yet are unrelated to this specific information intervention channel. In the appendix, we report impacts on students self-reports about skipping school, the importance of having a high grade point average, and an indicator for caring about higher grades more than their peers. Across all of these outcomes, we find treatment effects that are small in magnitude and not statistically significant.

5 Conclusion

This paper examines the network origins of pluralistic ignorance, focusing on the context of peer pressure about study behavior in American middle schools. In our

model, mistaken beliefs about what others think arises from the friendship paradox—people neglect the selection embedded in the neighborhood of peers they directly observe, producing biased estimates of the overall population. Our framework applies more broadly to settings where traits determining network formation correlate—or at least, are perceived to correlate—with traits about which people are trying to learn, and microfounds why misperceptions about others' beliefs are so often asymmetric.

References

Alesina, Alberto, Stefanie Stantcheva, and Armando Miano, "Immigration and Redistribution," *Review of Economic Studies*, 2022.

Allen, Franklin, Stephen Morris, and Hyun Song Shin, "Beauty contests and iterated expectations in asset markets," *The Review of Financial Studies*, 2006, *19* (3), 719–752.

Arteaga, Felipe, Adam Kapor, Christopher A. Neilson, and Seth Zimmerman, "Smart Matching Platforms and Heterogeneous Beliefs in Centralized School Choice," *Quarterly Journal of Economics*, 2022.

Banerjee, Abhijit, Emily Breza, Arun G Chandrasekhar, and Benjamin Golub, "When Less Is More: Experimental Evidence on Information Delivery During India's Demonetisation," *Review of Economic Studies*, 2024, *91* (4), 1884–1922.

Bernhadt, Arielle, Erica Field, Rohini Pande, Natalia Rigol, Simone Schaner, and Charity Troyer-Moore, "Male Social Status and Women's Work," *AEA Papers and Proceedings*, 2018, *108*, 363–367.

Braghieri, Luca, "Political Correctness, Social Image, and Information Transmission," *Working Paper*, 2021.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin, "Peer effects in networks: A survey," *Annual Review of Economics*, 2020, *12* (1), 603–629.

Breza, Emily, Arun Chandrasekhar, Tyler McCormick, and Mengjie Pan, "Using Aggregated Relational Data to Feasibly Identify Network Structure without Network Data," *American Economic Review*, 2020.

_ , Arun G. Chandrasekhar, Shane Lubold, Tyler H. McCormick, and Mengjie Pan, "Consistently estimating network statistics using aggregated relational data," *Proceedings of the National Academy of Sciences*, 2023, *120* (21), e2207185120.

Bursztyn, Leonardo, Alessandra González, and David Yanagizawa-Drott, "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia," *American Economic Review*, 2020, *110* (10), 2997–3029.

_ **and David Y Yang**, "Misperceptions about others," *Annual Review of Economics*, 2022, *14* (1), 425–452.

_ **and Robert Jensen**, "How Does Peer Pressure Affect Educational Investments?," *The Quarterly Journal of Economics*, 2015, *130* (3), 1329–1367.

_ **and** _ , "Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure," *Annual Review of Economics*, 2017, *9* (1), 131–153.

_ **, Georgy Egorov, and Robert Jensen**, "Cool to be Smart or Smart to be Cool? Understanding Peer Pressure in Education," *The Review of Economic Studies*, 2019, *86* (4), 1487–1526.

_ **,** _ **, and Stefano Fiorin**, "From Extreme to Mainstream: The Erosion of Social Norms," *The American Economic Review*, 2020, *110* (11), 3522–3548.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou, "Peer effects and social networks in education," *The review of economic studies*, 2009, *76* (4), 1239–1267.

Cantoni, Davide, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang, "Protests as Strategic Games: Experimental Evidence from Hong Kong's Anti-Authoritarian Movement," *Quarterly Journal of Economics*, 2019, *134* (2).

Conlon, John J, "Major Malfunction: A Field Experiment Correcting Undergraduates' Beliefs about Salaries," *Journal of Human Resources*, 2020.

Conlon, John J. and Dev Patel, "What Jobs Come to Mind? Stereotypes about Fields of Study," *Working Paper*, 2022.

DeGroot, Morris H, "Reaching a consensus," *Journal of the American Statistical association*, 1974, *69* (345), 118–121.

Delavande, Adeline and Basit Zafar, "Information and Anti-American Attitudes," *Journal of Economic Behavior and Organization*, 2018.

DeMarzo, Peter M, Dimitri Vayanos, and Jeffrey Zwiebel, "Persuasion bias, social influence, and unidimensional opinions," *The Quarterly journal of economics*, 2003, *118* (3), 909–968.

Erdös, P and T Gallai, "Gráfok elöirt fokú pontokkal," *Mat. lapok*, 1960, *11*, 264–274.

Friedkin, Noah E and Eugene C Johnsen, "Social influence and opinions," *Journal of mathematical sociology*, 1990, *15* (3-4), 193–206.

Fryer, Roland G. and Paul Torelli, "An Empirical Analysis of 'Acting White'," *Journal of Public Economics*, 2010, *95* (5), 380–396.

Golub, Benjamin and Matthew O Jackson, "Naive learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, 2010, *2* (1), 112–149.

Haaland, Ingar and Chris Roth, "Beliefs about Racial Discrimination and Support for Pro-Black Policies," *Review of Economics and Statistics*, Forthcoming.

Hastings, Justine, Christopher A. Neilson, and Seth D. Zimmerman, "The Effects of Earnings Disclosure on College Enrollment Decisions," *NBER Working Paper*, 2015.

Hayes, Alex, Arun Chandrasekhar, Tyler McCormick, and Emily Breza, "Spectral Estimation and Trait Selection for Aggregated Relational Data," *Working Paper*, 2026.

Jackson, Matthew O., "The Friendship Paradox and Systematic Biases in Perceptions and Social Norms," *Journal of Political Economy*, 2019, *127* (2), 465—972.

Jensen, Robert, "The (perceived) returns to education and the demand for schooling," *The Quarterly Journal of Economics*, 2010, *125* (2), 515–548.

Reeves, Steven Wilkins, Shane Lubold, Arun G Chandrasekhar, and Tyler H McCormick, "Model-based inference and experimental design for interference using partial network data," *arXiv preprint arXiv:2406.11940*, 2024.

Tankard, Margaret E. and Elizabeth Levy Paluck, "Norm Perception as a Vehicle for Social Change," *Social Issues and Policy Review*, 2016, *10* (1), 181–211.

Ye, Karen J., "Understanding Peer Effects in Educational Decisions: Theory and Evidence from a Field Experiment," *Working Paper*, 2022.

Zafar, Basit and Matthew Wiswall, "How do College Students Respond to Public Information about Earnings?," *Journal of Human Capital*, 2015.

6 Figures and Tables

Figure 1: Perceived Relationship Between Popularity and Study Effort

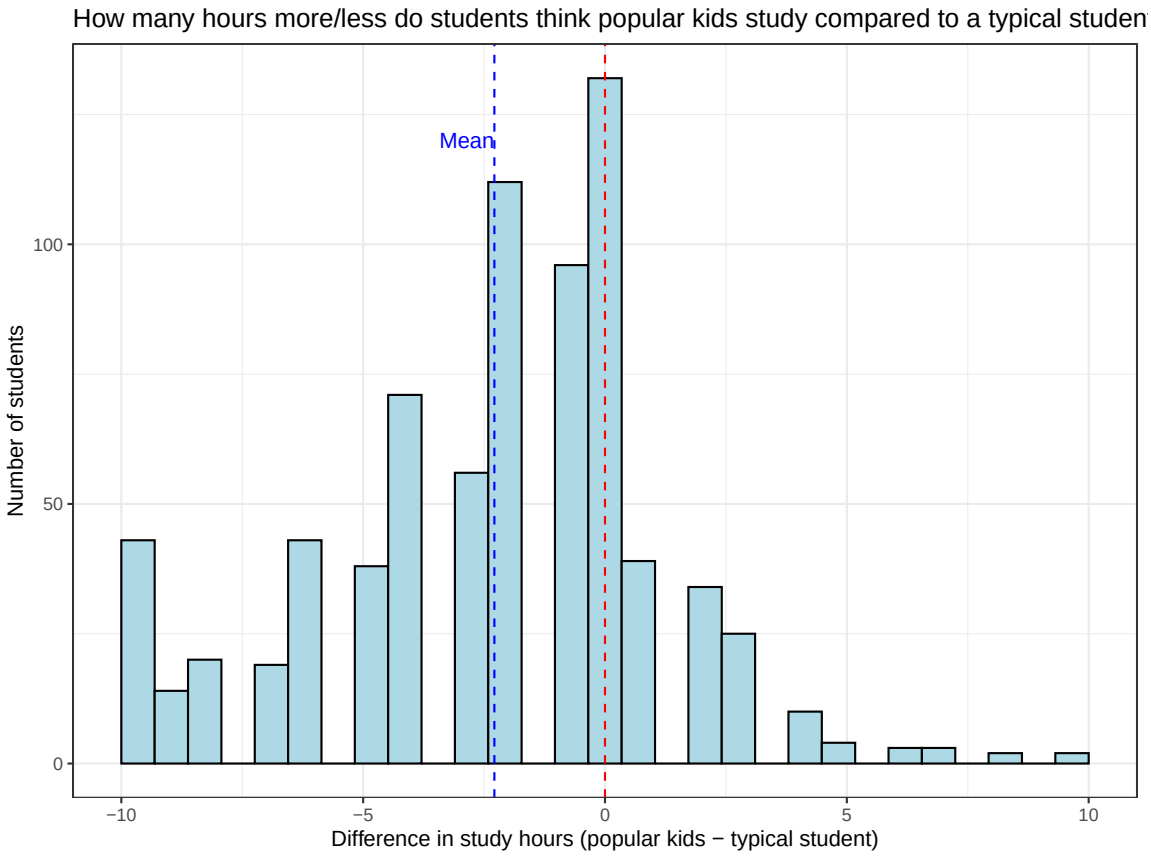


Table 1: Prior Belief about School-Wide Stigma and Average Neighbor Degree

	SBM		Spectral ARD		MCMC	
	(1)	(2)	(3)	(4)	(5)	(6)
Survey Degree		1.166**		1.343**		1.631***
		(0.442)		(0.561)		(0.491)
Average Neighbor Degree	2.159*	1.496**	1.591	1.518	-13.101	15.455
	(1.066)	(0.684)	(1.203)	(1.194)	(28.055)	(28.986)
Observations	363	363	328	328	331	331
R2	0.034	0.054	0.031	0.053	0.031	0.054
Outcome Mean	44.777	44.777	45.470	45.470	45.408	45.408
School-Grades	14	14	14	14	14	14
Fixed Effects	School-grade	School-grade	School-grade	School-grade	School-grade	School-grade
Controls	No	No	No	No	No	No

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Columns are grouped by ARD network method. Within each method, the first column regresses the outcome on average neighbor degree; the second adds survey degree. Survey degree is measured within grade as the number of named same-grade friends in the ARD survey. All regressions include school-grade fixed effects, cluster standard errors by school-grade, and include no additional controls.

Figure 2: Student's perception about relationship between study effort and popularity (Control sample only)

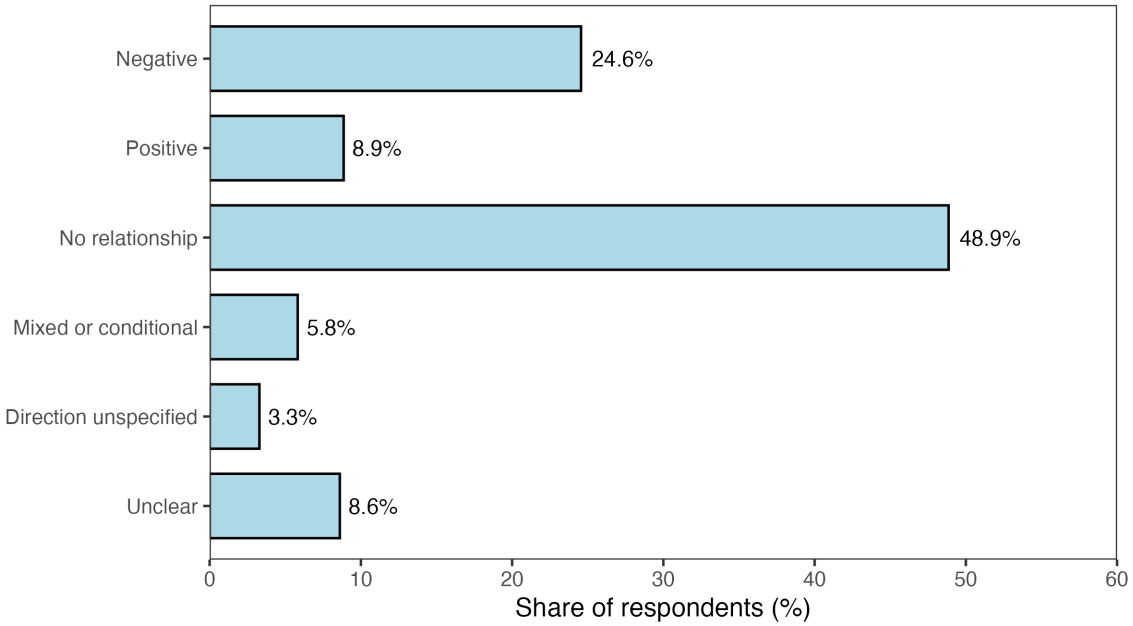


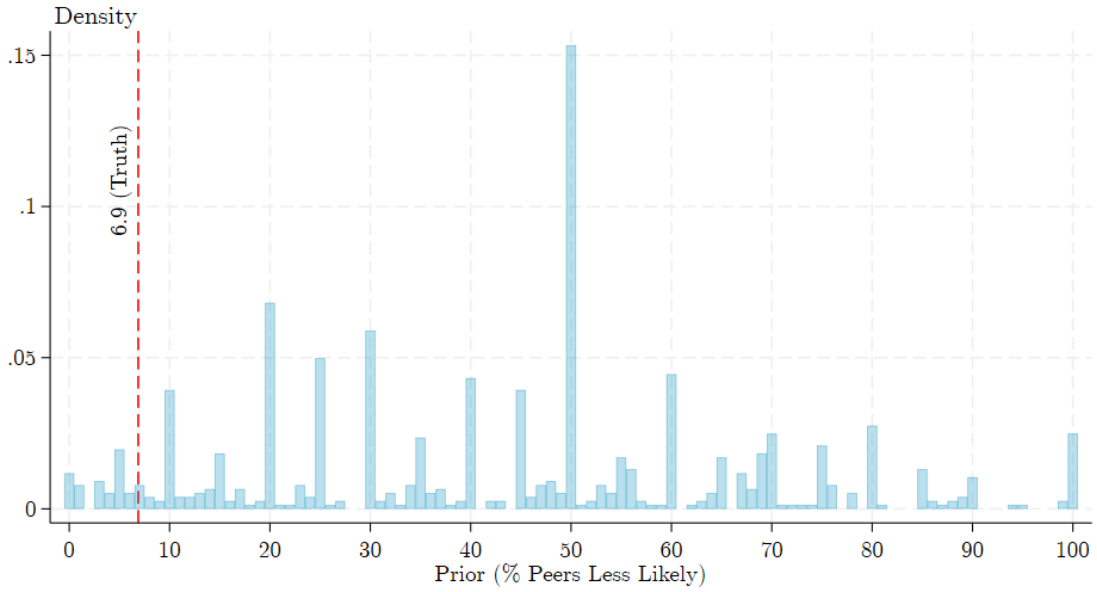
Table 2: Underestimation of Named-Friend Study Hours and Average Neighbor Degree

	SBM		Spectral ARD		MCMC	
	(1)	(2)	(3)	(4)	(5)	(6)
Survey Degree		-0.001		0.069		0.093
		(0.058)		(0.055)		(0.069)
Average Neighbor Degree	0.331**	0.333**	0.348***	0.344***	0.776	2.399
	(0.115)	(0.133)	(0.105)	(0.104)	(1.640)	(1.971)
Observations	336	336	330	330	333	333
R2	0.130	0.130	0.144	0.149	0.142	0.148
Outcome Mean	0.169	0.169	0.235	0.235	0.343	0.343
School-Grades	14	14	14	14	14	14
Fixed Effects	School-grade	School-grade	School-grade	School-grade	School-grade	School-grade
Controls	No	No	No	No	No	No

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Columns are grouped by ARD network method. Within each method, the first column regresses the outcome on average neighbor degree; the second adds survey degree. Survey degree is measured within grade as the number of named same-grade friends in the ARD survey. All regressions include school-grade fixed effects, cluster standard errors by school-grade, and include no additional controls.

Figure 3: Prior Peer Pressure Belief Compared to Truth



Note: Figure 3 shows the distribution of students' beliefs about negative peer pressure among their peers. The vertical red line denotes the truth. The first-order attitude question asks students about a "typical student" at their school: If you found out this student studied two hours more per week than you thought, how would that impact how much you would want to be friends with them? Everything else you know about the student is still the same. For example, they still have the same amount of time to spend with friends and faily or to play sports. Students who answer "much less" or "slightly less likely want to be their friend" are coded as negative peer pressure. The prior plotted in the histogram shows the distribution of second-order beliefs, which were elicited via the following question: Students taking this survey could have answered that they would be "slightly less likely" or "much less likely" to want to be friends with someone who studies two more hours per week. For every 100 students we surveyed, how many students do you think answered the question with either of these options? There are 100 boxes below to help you visualize your answer. When you type in an answer, the corresponding number of boxes will turn blue. The number of students out of every 100 can also be thought of as the "percent" of students answering this way.

Table 3: Typical Student Study Belief and Average Neighbor Degree

	SBM		Spectral ARD		MCMC	
	(1)	(2)	(3)	(4)	(5)	(6)
Survey Degree		0.003		-0.040		-0.079
		(0.035)		(0.040)		(0.049)
Average Neighbor Degree	-0.145**	-0.147***	-0.294*	-0.292*	-1.395	-2.777**
	(0.051)	(0.042)	(0.148)	(0.152)	(1.148)	(1.182)
Observations	366	366	329	329	332	332
R2	0.043	0.043	0.052	0.054	0.048	0.051
Outcome Mean	4.107	4.107	4.137	4.137	4.117	4.117
School-Grades	14	14	14	14	14	14
Fixed Effects	School-grade	School-grade	School-grade	School-grade	School-grade	School-grade
Controls	No	No	No	No	No	No

* p < 0.1, ** p < 0.05, *** p < 0.01
Columns are grouped by ARD network method. Within each method, the first column regresses the outcome on average neighbor degree; the second adds survey degree. Survey degree is measured within grade as the number of named same-grade friends in the ARD survey. All regressions include school-grade fixed effects, cluster standard errors by school-grade, and include no additional controls.

Table 4: Dependent Variable: WTP to Hide Grades

	(1)	(2)	(3)	(4)
Treatment	-7.297***	-12.767***	-6.461***	-7.213
	(2.327)	(2.967)	(2.409)	(4.481)
Treatment x (Prior - Info)		0.162**		0.021
		(0.071)		(0.121)
Observations	772	767	763	763
R2	0.010	0.015	0.042	0.042
Control Mean	17.951	17.951	18.130	18.130
Fixed Effects	No	No	Prior bin	Prior bin
Controls	No	No	No	No

* p < 0.1, ** p < 0.05, *** p < 0.01
Standard errors clustered by school-grade. Columns (3) and (4) absorb prior bins of width 10. No additional demographic controls included. No simulated ARD network statistics are used.

A Additional Figures & Tables

Table A.1: Binary Underestimation of Named-Friend Study Hours and Average Neighbor Degree

	SBM		Spectral ARD		MCMC	
	(1)	(2)	(3)	(4)	(5)	(6)
Survey Degree		0.003		0.006		-0.002
		(0.011)		(0.011)		(0.011)
Average Neighbor Degree	0.047**	0.043*	0.059***	0.059***	-0.136	-0.162
	(0.018)	(0.024)	(0.018)	(0.018)	(0.310)	(0.376)
Observations	336	336	330	330	333	333
R2	0.049	0.049	0.032	0.033	0.038	0.039
Outcome Mean	0.598	0.598	0.588	0.588	0.610	0.610
School-Grades	14	14	14	14	14	14
Fixed Effects	School-grade	School-grade	School-grade	School-grade	School-grade	School-grade
Controls	No	No	No	No	No	No

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Columns are grouped by ARD network method. Within each method, the first column regresses the outcome on average neighbor degree; the second adds survey degree. Survey degree is measured within grade as the number of named same-grade friends in the ARD survey. All regressions include school-grade fixed effects, cluster standard errors by school-grade, and include no additional controls.

Figure A.1: Error distribution about Typical Students' study hours

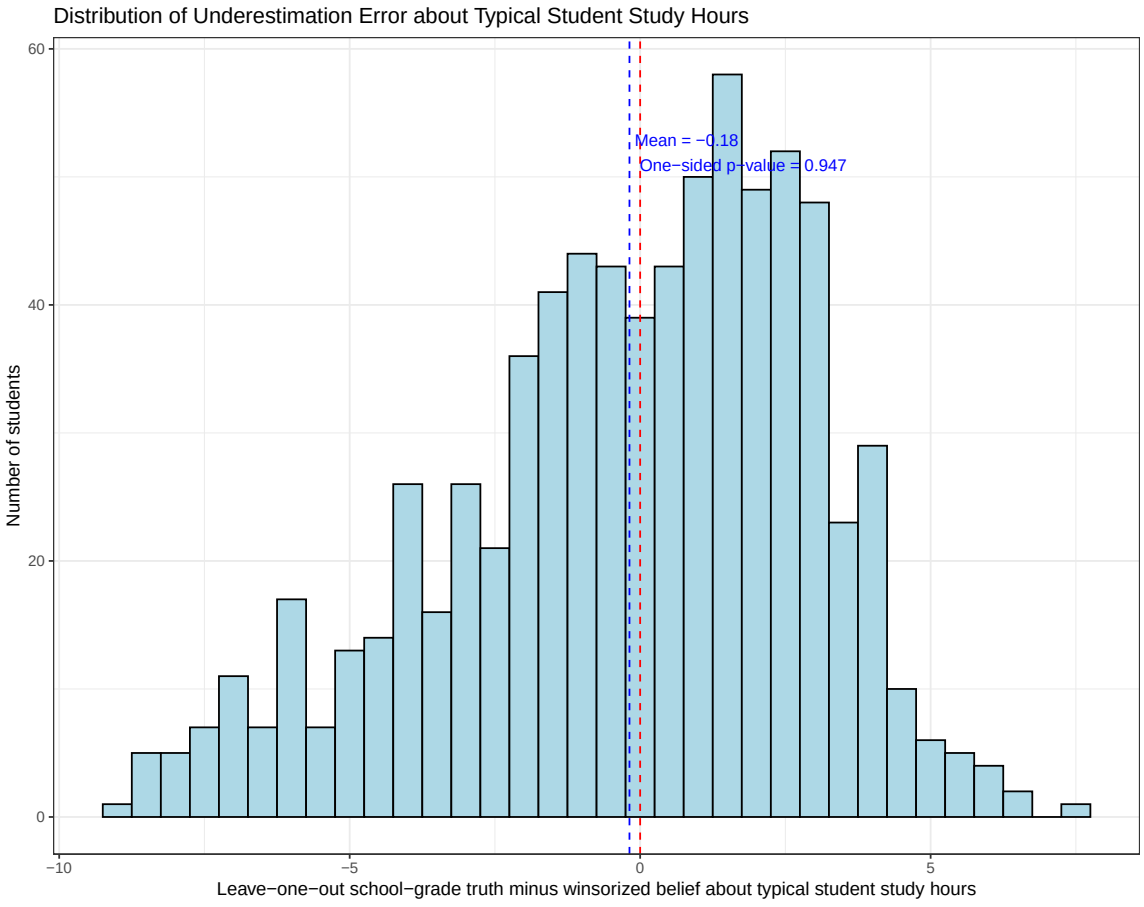


Figure A.2: Error distribution about Typical Students' study hours

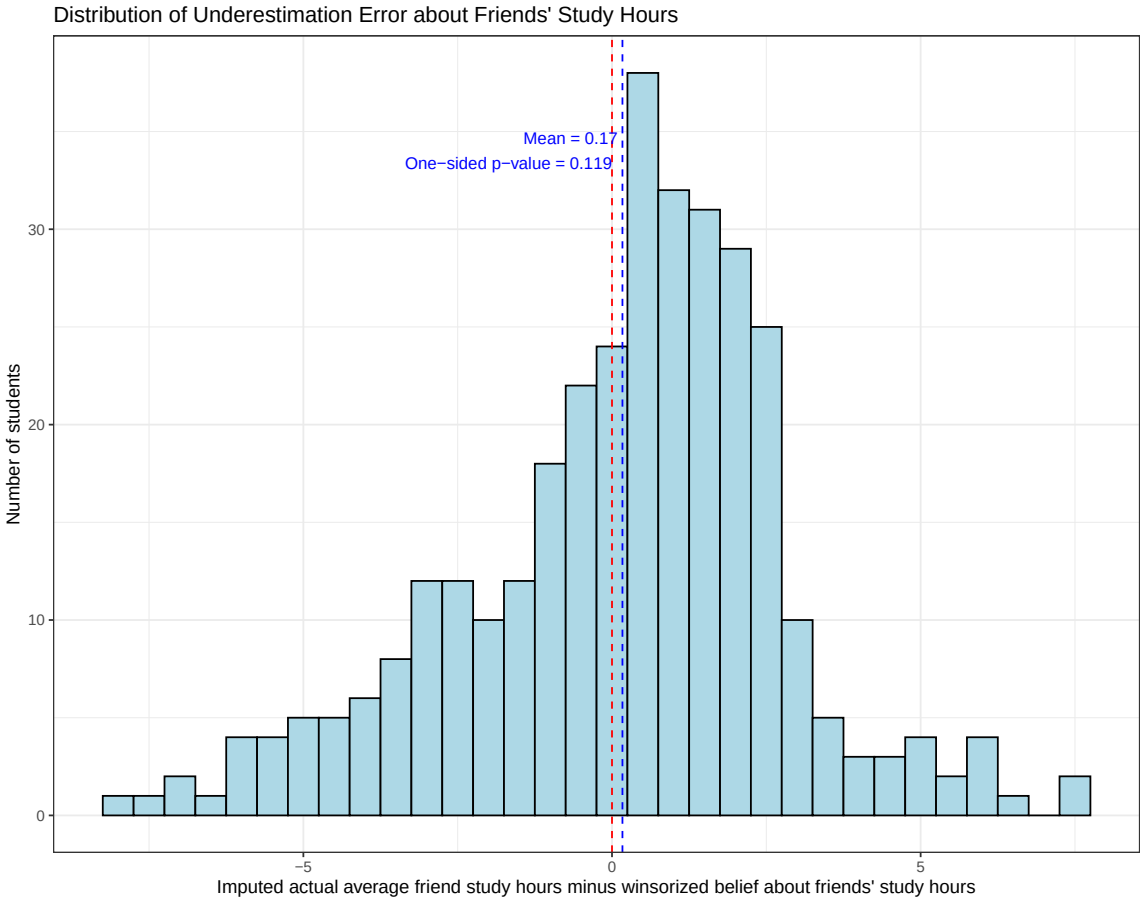
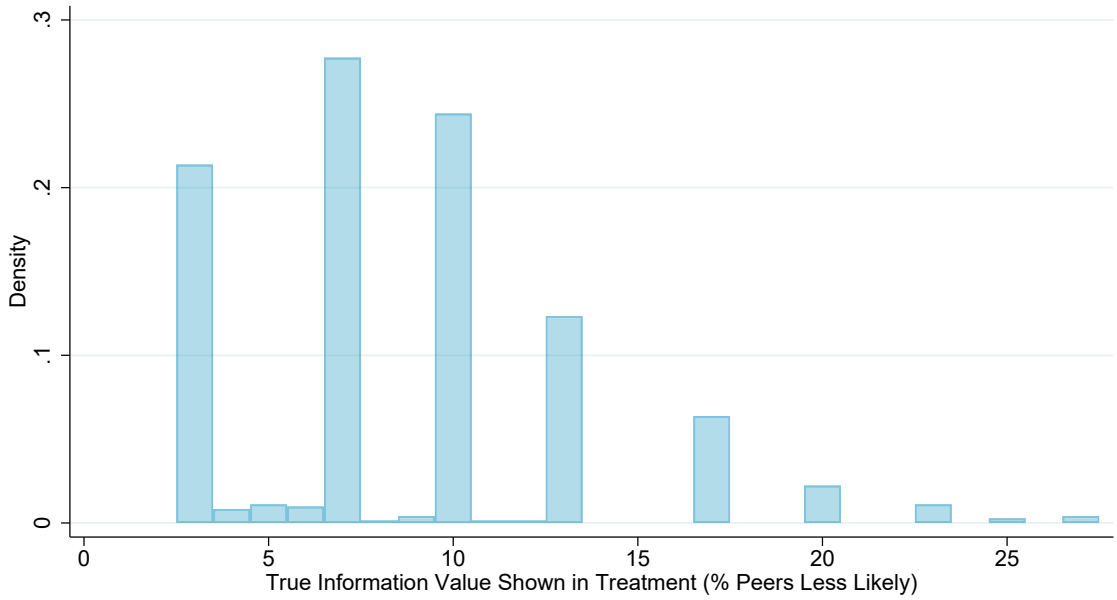


Table A.2: Overestimation of Stigma Across Demographic Groups

Share Overestimating	Size of Overestimation	Group	Sample Size
95.2	40.3	Female	419
92.7	37.9	Male	344
95.3	39.6	12 and under	493
91.9	38.6	Over 12	270
96.1	39.9	FRPL	179
86.4	45.6	ELL	59
94.6	40.9	Hispanic	280
93.9	39.4	White	574
93.5	37.9	Black	92
96.7	38.8	Multirace	30
94.1	39.2	Whole sample	763

Note: Column (1) of Table A.1 shows the share of respondents overestimating the stigma separately by demographic group. Column (2) shows the size of this overestimation conditional among sample of students who overestimate.

Figure A.3: Information Treatment Values



Note: Figure A.3 shows the distribution of values shown to the treatment group as part of the information treatment to correct their beliefs about negative peer pressure.

Table A.3: Survey Degree, Study Hours, and GPA

	Own Study Hours	Overall GPA
Survey Degree	0.045	0.053
	(0.049)	(0.076)
Observations	761	762
R2	0.122	0.121
Outcome Mean	4.004	89.690

* p <0.1, ** p <0.05, *** p <0.01
Standard errors clustered by school-grade. All specifications include school-grade fixed effects and no controls. Survey degree is measured within grade as the number of named same-grade friends in the ARD survey. No estimated or simulated network statistics are used.

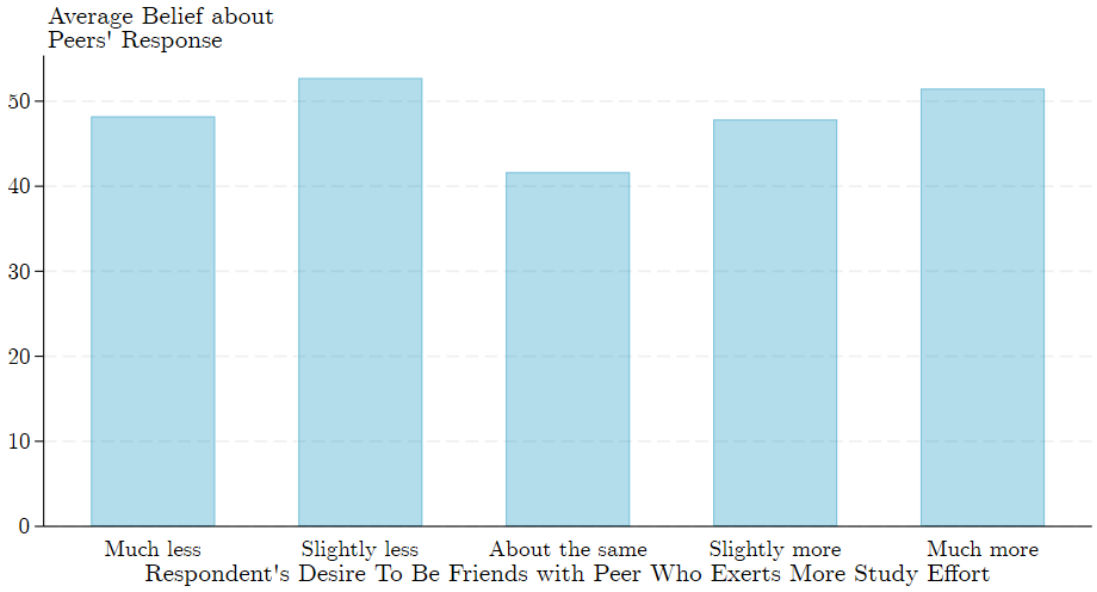
Table A.4: Own Stigma, Study Hours, and GPA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Own Study Hours	-0.012*						
	(0.007)						
Overall GPA		-0.008**					
		(0.003)					
Core GPA			-0.006**				
			(0.003)				
ELA GPA				-0.008***			
				(0.003)			
Math GPA					-0.004*		
					(0.002)		
Science GPA						-0.004	
						(0.002)	
Social Studies GPA							-0.003
							(0.002)
Observations	760	761	761	759	760	759	760
R2	0.054	0.055	0.055	0.061	0.052	0.052	0.052
Outcome Mean	-0.136	-0.135	-0.135	-0.136	-0.136	-0.136	-0.136

* p <0.1, ** p <0.05, *** p <0.01

Standard errors clustered by school-grade. All specifications include school-grade fixed effects and no additional controls. Higher values of own stigma mean greater stigma toward studying.

Figure A.4: Perceived Norm by Own Peer Pressure Attitude



Note: Figure A.4 plots the average second-order belief about negative peer pressure based on students' own first-order attitude. Negative peer pressure corresponds to the options "much less" and "slightly less". The first-order attitude question asks students about a "typical student" at their school: If you found out this student studied two hours more per week than you thought, how would that impact how much you would want to be friends with them? Everything else you know about the student is still the same. For example, they still have the same amount of time to spend with friends and faily or to play sports. Students who answer "much less" or "slightly less likely want to be their friend" are coded as negative peer pressure. The bars show the average of second-order beliefs, which were elicited via the following question: Students taking this survey could have answered that they would be "slightly less likely" or "much less likely" to want to be friends with someone who studies two more hours per week. For every 100 students we surveyed, how many students do you think answered the question with either of these options? There are 100 boxes below to help you visualize your answer. When you type in an answer, the corresponding number of boxes will turn blue. The number of students out of every 100 can also be thought of as the "percent" of students answering this way.

Table A.5: Experimenter Demand Placebo Test: RCT Impacts on Skipping School

	(1)	(2)	(3)	(4)
Treatment	-0.00110	0.0113	-0.000424	0.000704
	(0.0144)	(0.0202)	(0.0144)	(0.0234)
Treatment × (Prior - Info)		-0.000367		-0.0000320
		(0.000344)		(0.000467)
Observations	772	772	760	760
Control Mean	0.041	0.041	0.039	0.039
Controls	No	No	Yes	Yes

Note: Table A.5 presents results from the information experiment. The outcome is whether the student self-reports to having skipped school. All specifications report heteroskedasticity-robust standard errors. Columns (3) and (4) include fixed effects for the student's prior in bins of width 10, fixed effects for GPA bins of width 10 on a 100 point scale, fixed effects for past study hours, and gender. Statistical significance is denoted at the 10 percent level by *, at the five percent level by **, and at the one percent level by ***.

Table A.6: Experimenter Demand Placebo Test: RCT Impacts on Importance of High GPA

	(1)	(2)	(3)	(4)
Treatment	-0.591	-1.894	-0.742	-1.721
	(1.273)	(1.863)	(1.183)	(2.081)
Treatment × (Prior - Info)		0.0386		0.0278
		(0.0374)		(0.0484)
Observations	770	770	758	758
Control Mean	86.703	86.703	86.790	86.790
Controls	No	No	Yes	Yes

Note: Table A.6 presents results from the information experiment. The outcome is how important good grades are to the student on a 100-point scale, where the survey question read: "How important is it to you to get good grades?" All specifications report heteroskedasticity-robust standard errors. Columns (3) and (4) include fixed effects for the student's prior in bins of width 10, fixed effects for GPA bins of width 10 on a 100 point scale, fixed effects for past study hours, and gender. Statistical significance is denoted at the 10 percent level by *, at the five percent level by **, and at the one percent level by ***.

Table A.7: Experimenter Demand Placebo Test: RCT Impacts on Caring about Higher Grades than Peers

	(1)	(2)	(3)	(4)
Treatment	0.618	-4.189	1.595	-3.759
	(2.626)	(3.790)	(2.661)	(4.741)
Treatment × (Prior - Info)		0.142*		0.152
		(0.0822)		(0.110)
Observations	770	770	758	758
Control Mean	46.812	46.812	47.041	47.041
Controls	No	No	Yes	Yes

Note: Table A.7 presents results from the information experiment. The outcome is how much students say they care about getting better grades than their friends on a 100-point scale, where the survey question read: "How much do you care about getting better grades than your friends?" All specifications report heteroskedasticity-robust standard errors. Columns (3) and (4) include fixed effects for the student's prior in bins of width 10, fixed effects for GPA bins of width 10 on a 100 point scale, fixed effects for past study hours, and gender. Statistical significance is denoted at the 10 percent level by *, at the five percent level by **, and at the one percent level by ***.

B Model Appendix

Proof of Proposition 1. Fix an arbitrary type profile $(\theta_1, \ldots, \theta_n)$ and a degree d_i for an agent i. We begin by showing that

$$\mathbb{E}_i[\hat{\mu}_{s,i} \mid (\theta_1, \ldots, \theta_n), d_i] \leq \frac{1}{n-1} \sum_{j \neq i} \theta_j, \tag{2}$$

and the inequality is strict if $d_i > 0$.

To show inequality (2), we show that for any i, j, and j' distinct, we have $\mathbb{P}[j \in N_i] > \mathbb{P}[j' \in N_i]$ if and only if $\theta_j < \theta_{j'}$. Consider any graphical sequence $(d_1, \ldots, d_n)$ with $d_j \geq d_{j'}$ and d_i taking the value fixed above. The probability of this sequence is equal to

$$\alpha \prod_{\ell \neq i} F_{t^*_{\ell 2}}(d_\ell),$$

where $\alpha > 0$ is a normalization factor and $t^*_{\ell 2}$ is the optimal action for an agent of type θ_ℓ. The normalization factor comes from accounting for the non-graphical sequences, which do not change the relative probabilities of graphical sequences.

So by the monotone likelihood property of the family of distributions $F_t(d)$, the probability of $(d_1, \ldots, d_n)$ is higher than the probability of the sequence $(d'_1, \ldots, d'_n)$ defined by $d'_j = d_{j'}, d'_{j'} = d_j$, and $d'_\ell = d_\ell$ for all other ℓ. This sequence swaps d_j and $d_{j'}$. Under the configuration model, we have

$$\begin{aligned}
\mathbb{P}[j \in N_i \mid (d_1, \ldots, d_n)] &= \mathbb{P}[j' \in N_i \mid (d'_1, \ldots, d'_n)] \\
&\geq \mathbb{P}[j \in N_i \mid (d'_1, \ldots, d'_n)] \\
&= \mathbb{P}[j' \in N_i \mid (d_1, \ldots, d_n)]
\end{aligned}$$

and the inequality is strict if $d_j > d_{j'}$ and $d_i > 0$. Since this holds for any graphical degree sequence with $d_j \geq d_{j'}$ and d_i taking the value fixed above, we can conclude from our coupling that $\mathbb{P}[j \in N_i] > \mathbb{P}[j' \in N_i]$ if and only if $\theta_j < \theta_{j'}$. This completes the proof of inequality (2).

To complete the proof of the proposition, we can take expectations of both sides of inequality (2) with respect to d_i and $(\theta_1, \ldots, \theta_n)$. The left-hand side is equal to $\mathbb{E}_i[\hat{\mu}_{s,i}]$ by the law of iterated expectations, while the expectation of the right-hand side is equal to μ. □

Proof of Proposition 2. We begin by verifying the formulas. We have

$$\hat{\mu}_{t+1} = \lambda W\theta + W(1-\lambda)\hat{\mu}_t$$

for each $t \geq 0$. Since $\hat{\mu}_1 = W\theta$, we can expand the beliefs as

$$\hat{\mu}_{t+1} = \lambda \sum_{k=0}^{t-1}((1-\lambda)W)^k W\theta + ((1-\lambda)W)^t W\theta.$$

The sum on the right-hand side converges and the final term vanishes because W is stochastic, so the desired expression for $\hat{\mu}_\infty$ follows from the Neumann series expansion. The expression for θ_∞ is immediate. □

C Data Appendix

C.1 Data Collection and Sample Construction

This study was part of a larger data collection effort that included a variety of studies designed by scientists affiliated with Character Lab Research Network (CLRN). CLRN simultaneously rolled out multiple independent studies, and students were randomized to one of the studies running in their school. This study was conducted on school computers during class time in participating schools over the course of a two- to three-week testing window. On a predetermined testing day, a teacher proctor at each school administered the CLRN research activities to students. To introduce the study, teachers read a script that explained to students that all research activities were part of an educational research initiative at their school, that participation was voluntary and they were not being graded, and that teachers would not see their answers. Teachers also instructed students to focus on their own computers and (if relevant) not to look at classmates' screens. Upon logging into the CLRN platform, all students first viewed an assent screen that reiterated this information and, in addition, explained that parents would not see their responses and that their names and any other unique identifying information would not be shared with researchers. Students who agreed to participate were then directed to the survey.

We received responses from 1,279 students. This section details the sequential exclusion of responses until we arrive at our final sample of 772 students. Of the

original responses, we dropped 5 students who appeared to take the survey multiple times and an additional 160 who completed less than 50 percent of the survey. We implemented three basic attention checks. The first asked students "Are you going to answer questions on this survey as well as you can?" We drop an additional 20 students who say no. Second, in the middle of the survey, we ask students to "Please type the letter k below if you have been paying attention during this survey." We drop an additional 27 students who fail this check. Finally, at the end of the survey we asked, "Did you try your best to answer the questions in this survey as well as you could?" We exclude an additional 75 students who do not say "Yes, I tried my best". Next, we construct a series of attention checks based on student responses to study hour questions. Our main study hour question had an upper limit of 20 hours—if students listed 20 or more, then they were prompted in an open response to give the specific number. We exclude students who give an answer less than 20 in this open response stage. That includes 7 students when asked about their own study hours, 8 when asked about others, 14 when asked about their planned study hours, and 27 when asked about study hours of their friends. Next, we exclude 9 responses of students who said in the ARD elicitation that they did not have any close friends to list yet still listed at least one friend. We drop 105 students who switch multiple times in the price list elicitation for willingness-to-pay to hide their grades, and an additional 45 students whom cannot be matched to the administrative data from the schools. Finally, we exclude 5 students who do not say they live in Florida when asked. This gives the final sample of 772 students.

Table C.1: Balance Checks Between Included and Excluded Observations

	Included Obs.	Excluded Obs.	Diff.	S.E.	N
Age	12.14	12.19	-0.05	(0.06)	1,198
Grade	7.00	6.99	0.01	(0.05)	1,198
Female	0.55	0.48	0.07**	(0.03)	1,198
F.R.P.L.	0.24	0.37	-0.13***	(0.03)	1,198
E.L.L.	0.08	0.10	-0.02	(0.02)	1,198
White	0.75	0.70	0.05*	(0.03)	1,198
Hispanic	0.37	0.48	-0.12***	(0.03)	1,198
Black	0.12	0.21	-0.09***	(0.02)	1,198
Multi-Race	0.04	0.04	0.00	(0.01)	1,198
Total Absences	1.93	2.50	-0.57***	(0.18)	1,198
G.P.A.	89.63	85.15	4.49***	(0.45)	1,197
Share Survey Completed	98.73	74.74	23.99***	(1.14)	1,285

Note: Table C.1 presents balance tests between the observations included in the main sample and those excluded for the reasons discussed in section . The abbreviations "F.R.P.L.", "E.L.L." and "G.P.A." stand for free and reduced price lunch, English language learning, and grade point average. All variables except "Share Survey Completed" come from the administrative data provided by the district and not our survey. Columns 2 and 3 show the mean values for the included and excluded observations, respectively. Column 4 includes the difference in these two values, where *** denotes a p-value less than 0.01, ** denotes a p-value less than 0.05, and * denotes a p-value less than 0.10. The final column denotes the total number of observations, which varies based on availability of the administrative data and includes the potential duplicate responses from the survey data.

C.2 Survey Instrument

[The following includes our complete survey instrument, which was administered online.]

1. Thank you for taking part in our survey! Please answer the following questions as honestly as possible. Your answers will be kept confidential and anonymous.

2. Are you going to answer questions on this survey as well as you can? [Yes/No/I'm not sure]

3. How many hours per week do you spend doing homework and studying outside of class? [Slider between 0 and 20+ hours]

4. How many hours per week do you think the typical student from your school spend doing homework and studying outside of class? Please give your best guess. [Slider between 0 and 20+ hours]

5. How confident are you in your answer to the previous question about other students' weekly study hours? [Completely confident/Fairly confident/Somewhat confident/Slightly confident/Not confident at all. I have no idea how much other students study]

6. How much do you think the typical student at your school knows about how much other students study? [Slider between "Students barely know how much others study" and "Students know a lot about how much others study"]

7. Think about a typical student at your school. If you found out this student studied two hours more per week than you thought, how would that impact how much you would want to be friends with them? Everything else you know about the student is still the same. For example, they still have the same amount of time to spend with friends and family or to play sports. [Much less likely want to be their friend/Slightly less likely to want to be their friend/About the same/Slightly more likely to want to be their friend/Much more likely to want to be their friend]"

8. Thank you! We asked the question you just saw to other students who are taking the same survey as you. These students are from your school and schools just like your school. We are now going to ask you to predict how these other students answered the question. It's okay if you do not know the answer for sure, just give your best guess.

9. Students taking this survey could have answered that they would be "slightly less likely" or "much less likely" to want to be friends with someone who studies

two more hours per week. For every 100 students we surveyed, how many students do you think answered the question with either of these options? There are 100 boxes below to help you visualize your answer. When you type in an answer, the corresponding number of boxes will turn blue. The number of students out of every 100 can also be thought of as the "percent" of students answering this way.

10. Think about the students in your school who are the most popular. How many hours more or less do you think they study than the average student in your school? [Slider labeled "Difference in study hours (per week)" between -10 and 10, with labels denoting "Popular kids study less"/"Popular kids study the same as everyone"/"Popular kids study more"]

[*Treatment only:*] 11. Thank you for predicting how other students from your school and schools just like yours answered. Now, we will show you how 30 randomly-selected students from that group really answered the question.

[*Treatment only:*] 12. You predicted that XX% of students (out of every 100) taking this survey said that they would be "slightly less likely" or "much less likely" to want to be friends with someone who studies more. In fact, YY% of the students selected said that. Please type that answer below to make sure you remember it. There are 100 boxes below to help you visualize your answer. When you type in an answer, the corresponding number of boxes above will turn blue. Did anything surprise you about this fact? Why do you think your guess was different than the number reported above?

[*Control only:*] 13. Do you think studying impacts a student's popularity? Why or why not? [Free response]

14. Thank you! Please type the letter "k" below if you have been paying attention during this survey.

15. In this section, we will ask you about your close friends at your school. By "close friends" we mean those students that you spend lots of time with outside of class (in person or virtually), you talk to/text often, and you feel know you very well. We'll ask you to list the names or nicknames of your close friends. Please note that no one will ever see the names you list—your responses will not be saved and there is no way for anyone to access them. We only ask you to list your close friends so that we can ask about them by name later in the survey.

16. In the fields below, please list the names or nicknames of your close friends at your school. You DO NOT have to list 10 names. Please list only those whom you honestly consider close friends. If you have more than 10 close friends, please only list 10 of them.

[*Cycling through each friend and the student themselves (language below shown for the latter):*]
We'd now like to ask you questions about yourself.
17. Do you speak a language other than English? [Yes/No/I don't know]
18. Do you play a musical instrument? [Yes/No/I don't know]
19. Do you take a bus to school? [Yes/No/I don't know]
20. Are you on a sports team? [Yes/No/I don't know]
21. Do you use TikTok? [Yes/No/I don't know]
22. Did you go to the most recent school dance? [Yes/No/I don't know]
23. Have you skipped any classes in the last month? [Yes/No/I don't know]

[*Friends were additionally asked:*] 14. Is XXX in the same grade as you? [Yes/No/I don't know]
15. How many hours per week do you think XXX spends doing homework and studying outside of class? [Slider between 0 and 20+]
16. You just said that you think XXX spends at least 20 hours per week doing homework and studying outside of class. Exactly how many hours per week do you think they spend doing homework and studying outside of class? Please type the answer below. [Only shown to those who entered 20 in the previous question].

17. Thank you for your responses so far. We're now going to ask you a few more questions about studying.

18. How important is it to you to get good grades? [Slider between "Not Very Important" and "Extremely Important"] How important is it to your teachers that you get good grades? [Slider between "Not Very Important" and "Extremely Important"] How important is it to your parents that you get good grades? [Slider between "Not Very Important" and "Extremely Important"] How important is it to your friends that you get good grades? [Slider between "Not Very Important" and "Extremely Important"]

19. How much do you care about getting better grades than your friends? [Slider between "I don't care at all" and "I care a lot"]

20. For the rest of the school year, how many hours per week do you expect to spend on homework and studying outside of class? [Slider between 0 and 20+ hours per week]

21. Would you ever try to hide how much you study from your friends? [Never/Occasionally/Sometimes/Most of the time/All of the time]

22. Imagine that your grades from your most recent report card will be revealed to the rest of your school unless you pay some money. For each of the following amounts, would you pay that amount to prevent your grades from being revealed to the rest of the school? [Yes/No for the following options: $5, $10, $20, $40, $80, $160]

23. In general, in your school, do you think it is "cool to be smart"? [Slider between "Not at all cool to be smart" and "Very cool to be smart"]

24. What is the highest level of education that any of your parents/guardians have completed? [Less than high school diploma/high school diploma or GED/Started college, but no college degree/Associate's (2-year college) degree/Bachelor's (4-year college) degree/Post-graduate degree (Master's, Ph.D., medical degree, law degree, or other professional degree)]

25. What is the occupation of the parent/guardian in your household who makes the most money? If multiple parents/guardians make similar incomes, you only need to provide one answer.

The next set of questions will ask you about whether you feel lonely.

26. How often do you feel that you lack companionship? [Hardly ever or never/Some of the time/Often]

27. How often do you feel left out? [Hardly ever or never/Some of the time/Often]

28. How often do you feel isolated from others? [Hardly ever or never/Some of the time/Often] 29. How often do you feel lonely? ["Never"/"Hardly ever"/"Occasionally"/"Some of the time"/"Often/always"]

We're now going to ask you some questions about your school. These are the last questions of the survey!

30. What state is your school in?

31. How would you rate your school for each of the following? Move the slider to the right if you think your school is very good for the given dimension and to the left if you think it is very bad. [Classes/Sports/School Spirit/School buildings/School neighborhood]

32. Were there any parts of this survey that you found confusing? What were they? Do you have any other thoughts to share about this survey?

33. Did you try your best to answer the questions in this survey as well as you could? [Yes, I tried my best/No, I did not try my best/I'm not sure]

C.3 Using Aggregated Relational Data to Estimate Social Networks

This section summarizes our network estimation procedure using aggregated relational data (ARD). We estimate networks separately within each school-grade and follow the approach outlines in Reeves et al. (2024).

For each student i and trait t, let X_{it}^* denote the reported number of peers known by i with trait t, and let N_t denote the school-grade prevalence of trait t from administrative records. The set of traits that we use involve eliciting from the students in our sample, *"how many other students do they know who have trait t"* where the trait set is: speaks another language, plays a musical instrument, takes the school bus, is on a sports team, uses TikTok, attended the most recent school dance, and skipped class in the last month.

The key object in the first stage is each student's normalized ARD profile, $X_{it}^\dagger = X_{it}^*/N_t$. This normalization removes mechanical scaling from trait prevalence so that the remaining variation captures differences in social exposure. We cluster students into K latent groups using these normalized ARD vectors, producing estimated memberships $\hat{k}_i \in \{1, \ldots, K\}$. Let $\hat{\Omega}_{kt} = \hat{N}_{kt}/N_t$ denote the estimated share of trait-t students assigned to latent group k, where $\hat{N}_{kt}$ is the corresponding count in the estimated partition.

The second stage uses ARD moments to recover cross-group linking probabilities. Let $P_{kk'}$ denote the probability of an undirected edge between latent groups k and k', and define the trait-by-group connection profile $\tilde{P}_{kt} = \Pr(G_{ij} = 1 \mid k_j = k,\ t_i = t)$. Under the SBM structure, these objects satisfy the linear system

$$\tilde{P} = \Omega P,$$

where $\Omega \in \mathbf{R}^{T \times K}$ collects trait composition by latent group. With full column rank in Ω (and sufficient trait variation, typically $T \geq K$), this system identifies P from ARD moments. We estimate $\tilde{P}$ from observed ARD counts and recover $\hat{P}$ either from the linear system as

$$\hat{P} = \left(\hat{\Omega}^\top \hat{\Omega}\right)^{-1} \hat{\Omega}^\top \hat{\tilde{P}},$$

or, equivalently, by constrained least squares imposing undirected-network restrictions:

$$\hat{P} = \arg \min_{0 \leq P \leq 1,\ P = P^\top} \sum_{i,t} \left(X_{it}^* - \sum_{k'} \hat{\Omega}_{k't} P_{k'\hat{k}_i} \right)^2.$$

The symmetry and box constraints ensure that estimated block-level links are feasible

probabilities for an undirected graph.