The Costs and Benefits of Integration: The Case of Social Learning

Ari Anisfeld
University of Arkansas

Pepi Pandiloski
Harvard University
Job Market Paper *

March 6, 2026

CLICK HERE FOR THE MOST RECENT VERSION

Abstract

We study how the integration of different groups affects social learning. We designed a social learning game and implemented it in ethnically segregated schools in Macedonia. In the game, 6 players (3 students of 2 distinct ethnicities) get noisy private signals and need to learn the truth with their peers. This game allows us to control the learning environment by randomizing three dimensions: the network on which players observe each other, private initial signals, and bias, which we manipulate at the group level by randomizing the number of high-bias players. We find nuanced effects of integration on social learning. First, we find that bias increases the costs of integration on social learning. In low-bias groups, students learn more in the integrated network than in the segregated network, but the same is not true in the high-bias groups. Second, we find that complementary information increases the benefits of integration on social learning. Specifically, integration leads to more learning than segregation only in the games in which one ethnicity has more correct initial signals than the other ethnicity. In structural estimation of bias, we find that students learn from co-ethnics twice as much as they learn from cross-ethnics, and that bias is correlated with perceptions of education. We close the paper with exercises that simulate social learning under a counterfactual policy of one integrated classroom per cohort.

*Corresponding author: predragpandiloski@gmail.com. Pepi would like to extend an immense gratitude to James Robinson, Steven Durlauf, Oeindrila Dube, and Leonardo Bursztyn for their patient mentorship and generous help. We also thank Anjali Adukia, Aditi Bhowmick, Emily Breza, Ozan Candogan, Arun Chandrasekhar, Josh Dean, Yana Gallen, Claire Fan, Rachel Glennester, Donovan Harvey, Eduardo Montero, Jose Miguel Pascual, Violeta Petroska-Beshka, Raul Sanchez de la Sierra, Daniel Sonnenstuhl, Bryce Steinberg, as well as seminar participants at the Harris School of Public Policy PhD Workshop, the University of Chicago Development Economics Workshop and to the participants at the many brown bag lunches who provided feedback and moral support over the years. This research is funded by the JPAL European Social Inclusion Initiative, The Pearson Institute, the Development Economics Center and the Becker Friedman Institute at the University of Chicago. We'd like to thank Luan Imeri, Safet Ballazhi, and Zana Jusufi-Osmani at the Center for Human Rights and Conflict Resolution for excellent project implementation in the field, Jasmina Mihailovska and Tea Todorova at M-Prospect for coordinating data collection, and to 20 enumerators and 21 session facilitators whose diligent work made this study possible. We would like to thank Michael Plunkett and Federico Dominguez Molina for software development assistance in building the lab game platform, and Alex Chen for incredible research assistance in the field. We thank Arben Taravari, Aslian Snopce, and Martin Andonovski at Gostivar Municipality for their openness and collaboration.

1 Introduction

Understanding the costs and benefits of integrating different groups is a first-order issue. One theme in the literature is that bias increases the costs of integration. Biased preferences, where individuals prefer that their own group members benefit more, can lead to aggregate costs such as lower productivity (Hjort, 2014), lower provision of public goods (Alesina and Ferrara, 2005, Miguel and Gugerty, 2005) or conflict (Besley and Persson, 2011, Michalopoulos and Papaioannou, 2016). Another theme in the literature is that complementarities between groups increase the benefits of integration. If groups have different skills or knowledge, they can specialize and reap benefits from trade (Montalvo and Reynal-Querol, 2021, Jha, 2013) or increase productivity in firms, cities, or states (Nathan, 2016, Ottaviano and Peri, 2006, Peri, 2012).

In contrast, social learning has received less attention, but it is potentially no less vital: voters persuade each other about policies, farmers rely on social learning when adopting new fertilizers, and high school students may pick up smoking because their peers are doing it. How does integration of different groups influence social learning? Conventional wisdom suggests that integration improves learning, since information can travel between groups and avoid echo chambers (e.g. Golub and Jackson, 2012, Chandrasekhar et al., 2020, Acemoglu et al., 2011). However, individuals may be biased when learning from other groups. Republicans may dismiss Democrats' beliefs, farmers of different castes may not follow each other's agricultural practices, and high schoolers of different ethnicities may imitate co-ethnics more.

In this paper, we find that integration has nuanced effects on social learning that depend on both bias and complementarities of prior information. Bias in learning increases the costs of integration: if individuals discount the actions of members of different groups, then integration weakly decreases social learning. On the other hand, complementary information increases the benefits of integration: if one group has more correct initial information than the other, then integration weakly increases learning.

We study this question in segregated high schools in Gostivar, a town in Macedonia with three dominant ethnic groups: Albanians, Macedonians, and Turks. While the country separated peacefully from Yugoslavia in 1991, an armed ethnic conflict between Macedonian authorities and Albanian separatists broke out in 2001, and was subsequently ended with the power-sharing Ohrid Agreement. As a result, students in Gostivar attend mixed-ethnicity schools, but are educated in classrooms segregated by language of instruction. Native

language instruction might be desirable: it gives cultural recognition to ethnic groups and can increase human capital. However, such parallel institutions have resulted in chronic segregation in the high schools we study: 55% of the students in our sample cannot name a single cross-ethnic friends, and students are 16 times more likely to greet a random co-ethnic than a random cross-ethnic.[1] These segregated schools are a particularly pertinent setting for our question. Social learning shapes a host of important outcomes among adolescents, including desirable behaviors like wearing covid-protecting masks as well as harmful behaviors like smoking. Moreover, schools function as sites of socialization that can mold young students into the citizens of a peaceful, cooperative, and prosperous future. To this goal, the Municipality is considering creating one integrated English language classroom per cohort, which motivates our question. Integration may increase the availability of (good) information, but bias may hinder learning, so the answer is not obvious.

How integration affects social learning is a difficult question to study empirically. First, observed relationships do not just capture avenues of social influence, but also common characteristics between individuals.[2] If two friends are doing the same thing, it is hard to know if this is due to social learning or individual characteristics, an issue known as the correlated effects problem.[3] Second, groups may have asymmetric preferences, such that a different "state of the world" is "true" for different groups. As a consequence, what may seem as an epistemic failure, may be a case of *de gustibus non est disputandum*.[4] To address these empirical issues, we combine a theoretical model with a lab-in-the-field experiment.

To generate testable hypotheses, we extend a commonly used model[5] of social learning on networks to allow for biased agents. In the network formation stage, *integration* grows as different-group links become more common relative to same-group links, holding constant the average number of links.[6] In the social learning stage, agents first get noisy random signals. The groups have *correlated priors* if one group has a more precise initial signal than the other.[7] Agents learn by playing a coordination game with their network neighbors. We

[1] In a network survey, students were given 4 randomly selected co-ethnic students and 4 randomly selected cross-ethnic students and were asked whether they would say hi to them if they passed each other in the school hallway. Students report saying hi to 55.3% of their coethnics and to 3.5% of their cross-ethnics

[2] For example, neighborhoods or classrooms may present opportunities to meet and form social ties. Individuals also tend to form relationships with those who are similar to them, both in characteristics and behaviors.

[3] See Bramoullé et al., 2020 on the correlated effects problem in the identification of peer effects, and common solutions to it.

[4] Tr. from Latin: "In matters of taste, there can be no disputes." For example, it would be a mistake to interpret the observation that Muslims don't eat pork as a failure of social learning from Christians; rather, Muslims and Christians have different religious prescriptions that imply different utility from eating pork.

[5] E.g. Golub and Jackson, 2012, Banerjee et al., 2021

[6] This holds constant the overall density of the network Intuitively, segregation arises from "rewiring links" rather than adding or removing links in the network, which have a mechanical effect under the model.

[7] We hold constant the initial true signals for the whole network (fixed overall precision), which can have change learning me-

model *bias* as agents' relative preference for coordinating with their own versus a different group. The model suggests that: 1) bias weakly increases the costs of integration, and 2) correlated priors weakly increase the benefits of integration.

To test these hypotheses, we implement a social learning game in ethnically segregated high schools in Macedonia. The game *"Red or Blue"* is played by a group of 6 students – 3 students from 2 distinct ethnicities. In this game there are two bags, a red bag (with 4 red balls and 2 blue balls) and a blue bag (with 4 blue balls and 2 red balls). One bag is randomly chosen and players have to guess the color of the chosen bag. At the start, each player randomly gets one of the 6 balls as a "private signal". Thus, 4 students get a correct signal and 2 students get an incorrect signal. In the first round, the players see their own ball and make a guess. In subsequent rounds, the players observe each other on a social network. Each player can see their own previous-round guess and the previous-round guesses of their network neighbors, after which they guess again.[8] If students pick the correct color in the final round of the game they get a cash reward.

This game allows us to induce experimental conditions that directly map onto the model parameters. First, we induce random integration by varying the observation networks during each game. Figure 1 visualizes the *Segregated* and *Integrated* treatments. Integration is varied by fixing the number of observed players, but changing their ethnic identity. Second, we vary the correlation of the priors (private signals). In the *Uncorrelated priors* arm both ethnicities have the same prior information: 2 players get a correct signal and 1 player gets an incorrect signal. In the *Correlated priors* arm, one ethnic group has better information than the other: one ethnicity gets 2 incorrect and 1 correct signal, while the second receives 3 correct signals. Finally, we manipulate bias by randomizing the types of players in the group of 6. Specifically, in our baseline survey, about half of the participants do not report any cross-ethnic friendships, which we use as a proxy for high-bias. As a consequence of different student types, some 6-person groups will be more biased by chance, and a median split helps us manipulate bias: *High bias* groups are those in which 3 or fewer students have a cross-ethnic friend, while the *Low bias* groups are those in which 4 or more students have a cross-ethnic friend. This design addresses the empirical problems associated with our question.[9]

chanically. Intuitively, it creates an environment in which one ethnicity can stand to learn from the other, while holding fixed the overall initial social knowledge.

[8] The computer shows the players' objective guesses, precluding cheating or other strategic communication.

[9] By randomizing the network structure and the initial signals, we deal with the correlated effects problem since we eliminate the possibility that any common characteristics drive initial actions and preclude endogenous selection of peers. Additionally, students from different ethnicities show similar preferences for red and blue, alleviating concerns about asymmetric preferences.

We find support for the hypothesis that bias increases the costs of integration. In low-bias groups, integrated networks lead to more learning than the segregated networks, but the same is not true in high-bias groups. Most importantly, we can reject the null hypothesis that bias decreases the cost of integration at the 10% significance level, leading us to conclude that our experimental data is consistent with the first hypothesis. This result does not have a causal interpretation, and we cannot conclude that making a friend makes one learn better in an integrated network. Rather, we interpret it as a heterogeneous effect of integration that depends on the bias of the individual. A limitation of our manipulation of bias is that it may be confounded. For example, if students with cross-ethnic friends are more trusting or have higher cognitive abilities, it may be that these drive our results instead. In exploratory analyses, we show that our result is robust when we use different bias proxies, and unlikely to be driven by general trustworthiness or cognitive skills.

We also find evidence for the second hypothesis that correlated priors increase the benefit of integration. In games with correlated priors, integration leads to more learning than segregation, while the same is not the case for games with uncorrelated priors. Again, we are able to reject the null hypothesis that correlated priors decrease the benefit of integration at the 10% significance level, leading us to conclude that the experimental data is consistent with the model. Our results imply nuanced effects of integration: it can be harmful if groups are sufficiently bias against one another, but beneficial if they stand to learn from one another. They suggest that in practice, it is this interaction of bias and complementarity of information that may ultimately decide the effect of integration.

The tight correspondence between the model and the game allows us to go beyond the lab in two ways. First, we fit the model to the experimental data to estimate the bias. We find that students are biased against everyone, but significantly more toward cross-ethnic students. Students follow their own actions twice as much as they follow the actions of co-ethnic players, but almost five times as much as they follow the actions of cross-ethnic players. In other words, students learn from their co-ethnics more than twice as much as they learn from their cross-ethnics. We correlate individual-level bias estimates with student characteristics, and find that the strongest correlation is between perceptions of education and bias – the players who perceive an ethnic group to be more educated also have lower bias toward that group during the game. Second, we simulate social learning under counterfactual policies. The Mayor of Gostivar has spoken about plans to open one integrated classroom per school cohort. We combine bias estimates with survey data on real-life

networks and simulate changes in social learning as a consequence of classroom integration.

Our study makes several core contributions. First, the political economy and development literature has long focused on the consequences of integrating different groups. On one hand, biased preferences can lead to social costs of integration, such as lower levels of public goods (Alesina and Ferrara, 2005, Alesina et al., 1999, Miguel and Gugerty, 2005, Banerjee et al., 2005, Khwaja, 2009) or intergroup conflict (Besley and Persson, 2011, Montalvo and Reynal-Querol, 2005, Esteban and Ray, 2011, Berman et al., 2017, Michalopoulos and Papaioannou, 2016). On the other hand, complementarities in skills or endowments can engender social benefits of integration, through specialization and trade (Montalvo and Reynal-Querol, 2021, Bahar and Rapoport, 2018, Jha, 2013, Alesina et al., 2016) or in joint production (Lazear, 1999, Alesina and Ferrara, 2005, Nathan, 2016, Ottaviano and Peri, 2006, Peri, 2012). We expand on this literature by bringing to light how integration shapes social learning, an outcome that has received little attention but is important. Interestingly, we find analogous mechanisms at work: bias increases the costs while complementarities increase the benefits of integration. This finding is theoretically generated by a model that is very different from the ones that generate the classical results in the political economy and development literature, and is then tested and confirmed by a field experiment.

Second, a large literature studies social learning across domains.[10] Our contribution to this literature is threefold. First, to our knowledge, we are the first to study how identity interacts with network structure, a question that has eluded empirical treatment.[11] It is important to study social learning on a network of individuals, as we can capture spillovers and group-level dynamics like echo chambers. Second, we find empirical support for a conventional wisdom from the theoretical literature on the benefits of integration (Golub and Jackson, 2012, Chandrasekhar et al., 2020). However, these models usually assume away bias, something that is central in our study. As a consequence, we introduce new results about the potential costs of integration that are driven by bias. Finally, we find that agents incorporate more information from co-

[10] Examples include the spread of agricultural technology (Conley and Udry, 2010, Foster and Rosenzweig, 1995), take up of insurance (Cai et al., 2015) and microfinance (Banerjee et al., 2013), financial decisions (Bursztyn et al., 2014), spread of health products (Kremer and Miguel, 2007, Dupas, 2014, Oster and Thornton, 2012), retirement savings (Duflo and Saez, 2002, Duflo and Saez, 2003), job referrals (Granovetter, 1983, L. A. Beaman, 2012, L. Beaman and Magruder, 2012, Goel and Lang, 2019, Bayer et al., 2008, Bayer et al., 2008, Damm, 2009), voting behavior (Lazarsfeld et al., 1944, Zelizer, 2019), and crime (Glaeser et al., 1996)

[11] This is largely due to the identification issues we raise in this paper. While the empirical literature has considered how identity might shape social learning (e.g. Munshi, 2004, Miller and Mobarak, n.d., Bayer et al., 2008, Conlon et al., n.d.), the question of how network structure interacts with identity has eluded empirical treatment. We are able to achieve this by maintaining a close relationship between a theoretical model and an empirical experiment, the lack of which has been lamented (Mobius and Rosenblat, 2014)

ethnics than cross-ethnics. The leading explanation for information discounting is an "ownership bias" effect (Conlon et al., 2022). While this can explain why co-ethnic peers are discounted, it cannot explain the additional discounting of cross-ethnic peers. Thus, our study suggests that people may discount information based on identity and raises future questions about how identity can lead to epistemic failures.

The paper proceeds as follows. The next section details the local context in Macedonia. Section 3 introduces the model and generates theoretical hypotheses. Section 4 details the study design and data. Section 5 tests the hypotheses. Section 6 furthers our empirical results by structurally estimating learning bias and presenting counterfactual policy simulations. Section 7 concludes.

2 Ethnicity and education in Macedonia

Our setting is Macedonia, a small but diverse country in the heart of the Balkan peninsula. Macedonians, Albanians, Turks, and other ethnic groups have lived here together for hundreds of years. Nationally, Macedonians are the majority, with about 60% of the population, followed by Albanians who compose about 25%, and Turks which come at 4%; however, in Gostivar Municipality, where we conducted our study, Albanians are the majority (55%), followed by a sizable minority of Macedonians (21%) and Turks (13%). Each of the three dominant ethnicities has a different language, all of which are mutually incomprehensible. Religion also divides the groups: most Macedonians are Orthodox Christian while most Albanians and Turks are Sunni Muslim.

There is no clear hierarchy between the three dominant ethnicities in Gostivar. Between the end of World War II and 1991, Macedonia was part of socialist Yugoslavia – "the land of the South Slavs". As Slavs, Macedonians enjoyed cultural and political privileges that other ethnicities did not, including better access to public goods and employment, and education in their native language at all levels. Disadvantaged in Yugoslavia, many Albanians emigrated to work in rich but relatively near countries such as Germany and Switzerland. Keeping close ties with their extended families in Gostivar, Albanian expats flock back to Macedonia every summer, a season of weddings, building new houses for the next generations, and reuniting with loved ones. The high levels of remittances from Albanians abroad to their families in Gostivar, as well as capital investments of the returning migrants, have ensured that Albanians have access to resources

despite their systematic exclusion in Yugoslavia. Finally, the Turkish population descends from the ruling class during the Ottoman Empire, which ruled in Gostivar for almost 5 centuries and until the end of the First Balkan War in 1913. Many of the Turks in Gostivar are members of the historical *esnaf* class – craftsman such as barbers, pastry chefs, shoemakers, and tailors – who have historically owned houses and stores in central locations in town. Customary apprenticeships among the *esnaf* Turks and centrally located property have ensured that Turks also have access to skills, jobs, and resources. With similar access to resources, each group has remained its own separate world, so neither economic nor social inter-dependencies are strong pushes toward integration.

Survey data on stereotypes among the three dominant groups suggests that each ethnicity has different views on social hierarchies. We asked the students in our sample questions from the Stereotype Content Model (Cuddy et al., 2008), which elicits attitudes about groups in two dimensions: *warmth* and *competence*. Figure 2 shows the results. All three ethnicities see their own group on top of the social hierarchy. Albanians and Turks are similar; they see the other two ethnicities in Gostivar as having lower warmth and competence, but not too different from one another. On the other hand, Macedonians see Turks as similar to themselves, while ranking Albanians as lower on competence and warmth. The picture that emerges from the data supports the view that each ethnicity has different perceptions of the social hierarchy and how groups fit in it. It seems that history has shaped some of these stereotypes: for example, both Macedonians and Albanians agree that Macedonians are more educated than Albanians, but that Albanians are richer than Macedonians.[12] This is not surprising given the historically higher access of Macedonians to education, and the high remittances from Albanian expats into Gostivar.

Exclusion of the large Albanian minority in Yugoslavia boiled over during the early years of independent Macedonia, the turbulent 1990s when the Balkans collapsed into war. Albanians demanded more rights and institutional inclusion, and education has always been at the forefront of the clashes. In 1995, Albanians opened an "illegal" university, which Macedonian authorities forcefully closed, killing one Albanian and injuring eleven in the process. In the following years, cycles of Albanian protests demanding more rights, and government suppression of these protests escalated into an armed conflict between Albanian militias and Macedonian forces in 2001. This armed conflict was resolved with the Ohrid Agreement, which granted Albanians more political and cultural rights, including quotas for public employments and education in their

[12]The results by separate dimensions are not shown.

native language at all levels.

The informal norms in the mixed schools in Gostivar have gone beyond policy and created completely parallel institutions. Mayors have settled on a political tradition that prescribes for the 8 mixed schools in the Municipality, 5 Albanian principals, 2 Macedonian, and 1 Turk – a very close replication to the relative size of the groups in the area. Furthermore, each mixed school has a vice principal that represents a different ethnicity than the one of the principal. In reality, these vice principals function as principals for their own ethnic group, a division that is replicated for almost all school employees. This power-sharing agreement has cemented a chronic segregation. Students go to the same school building, but are all segregated into different classrooms, which instruct in the three local languages. Students have fixed classrooms, such that different subject teachers rotate to a fixed group of students in a fixed room. Often, different ethnic groups are organized in different shifts, and even when they are integrated within the same shift, often have different wings in the schools assigned to them. Our baseline network survey shows the extent of segregation. About 55% of the students in our sample cannot name a single cross-ethnic friends, and the students are more than 16 times more likely to greet a random co-ethnic than to a random cross-ethnic in their cohort. During a network survey, students were given 4 randomly selected co-ethnic students and 4 randomly selected cross-ethnic students from their school cohort, and were asked whether they would say hi to that student if they passed each other in the hallway. On average, students report greeting 55.3% of their named co-ethnics and 3.5% of their named cross-ethnics.

These high schools in Macedonia are particularly well suited to study how integration of different groups affects social learning. While the history of the ethnic groups is unique in its own right, Macedonia is similar to many other developing countries with linguistic diversity and weak states, where different ethnic groups have not united under a common political, ideological, or economic project, but live in parallel universes alongside each other. The schools are also an attractive feature, partly because they replicate the social divisions, but partly because of the age of the students. Adolescents in high school rely on peers for a lot of important decisions and social learning among them can shape a host of important outcomes, some good like wearing covid-protecting masks, some bad like smoking. This malleability of the future citizens make schools a crucial site of socialization, one that can ensure a more peaceful and prosperous future among the different ethnic groups. In this spirit, the Mayor of Gostivar has proposed a policy of creating integrated

English-language classrooms in the high schools, and that they are negotiating this policy with the Ministry. If one classroom in each cohort is integrated, how might this change social learning? In Section 6.2 we conduct counterfactual exercises to answer this question, using the model in the following section.

3 Integration and social learning with biased agents

3.1 Model

In this section, we present our model and theoretical hypotheses. In the network formation stage, agents make links with probabilities that depend on the groups agents belong to. In the social learning stage, agents get noisy signals about a binary state of the world, and play a myopic coordination game on the network. This framework allows us to formalize integration, correlated priors, and bias in learning, and generate hypotheses about the effects of integration.

Network formation

Consider a population of agents $N = \{1, 2, ..., n\}$, such that each individual i has one of two types: $k_i \in \{1, 2\}$. Each type is represented equally in the population.[13] Each possible pair ij meets once, and at this meeting, they form a link with probability:

$$\mathbb{P}(ij \text{ form a link }) = \begin{cases} 1 & \text{if } i = j \\ p_s & \text{if } i \neq j, k_i = k_j \\ p_d & \text{if } i \neq j, k_i \neq k_j \end{cases} \tag{1}$$

Since we want to explore consequences of integration, we focus on networks in which links within each group are at least as likely as links across groups: $1 \geq p_s \geq p_d \geq 0$. We define **integration** as the ratio $\frac{p_d}{p_s} = I \in [0, 1]$, holding the expected number of links $p_s(\frac{n}{2} - 1) + p_d\frac{n}{2}$ fixed in expectation. Intuitively, fixing $p_s(\frac{n}{2} - 1) + p_d\frac{n}{2}$ corresponds to a "rewiring" of the network, rather than adding or removing links.[14]

[13] We take n to be even, and focus on the 2 group case to most simply convey the intuitions.

[14] We do not consider a model in which integration increases if we add more cross-group links. In this case, the model predicts weakly more learning because there is more access to information, which is a trivial and uninteresting case. Such case also conflates

This process results in a social network described by an $n \times n$ symmetric adjacency matrix $\mathbf{G}$ with a characteristic entry:

$$G_{ij} = \begin{cases} 0 & \text{if } ij \text{ are not linked} \\ 1 & \text{if } ij \text{ are linked and } i = j \\ 1 - b_{si} & \text{if } ij \text{ are linked, } i \neq j, \text{ and } k_i = k_j \\ 1 - b_{di} & \text{if } ij \text{ are linked, } i \neq j, \text{ and } k_i \neq k_j \end{cases} \tag{2}$$

The parameters b_{si}, b_{di} capture the learning bias by agent i toward individuals of the same and different groups respectively, and reflect that social network ties can be weaker or stronger – the larger the bias the weaker the tie. Agents are heterogeneous and bias parameters are randomly and independently drawn from two normal distributions, i.e. $b_{si} \sim N(b_s, \sigma_s), b_{di} \sim N(b_d, \sigma_d)$. Thus, the parameters $b_i = (b_{si}, b_{di})$ define a type who treats information from all individuals of a given group equivalently. This bias will be important in the social learning stage later.[15]

Social learning

The social learning stage unfolds over discrete time. At $t = 0$, each agent i receives a private binary signal $\omega_i \in \{-1, 1\}$ such that the probability it is correct depends on the agent's type:

$$\mathbb{P}(\omega_i \text{ is correct}) = \begin{cases} s_1 & \text{if } k_i = 1 \\ s_2 & \text{if } k_i = 2 \end{cases} \tag{3}$$

with $s_1, s_2 \in [0, 1]$. While the precision of the signal can be different by group, most individuals in the network receive the correct signal, i.e. $\frac{s_1+s_2}{2} > 0.5$. A group with a weaker signal can learn from the other group. For example, if group 1 has a stronger signal than group 2, i.e. $s_1 > s_2$, then the individuals of group 2 can learn more from the individuals of group 1, since they have better information on average. We call such

measurements of integration with overall density in the network.

[15]The network formation model is often referred to as the stochastic block model and belongs to one of the most widespread family of models in the literature (see Chandrasekhar, 2016 for a review of stochastic block models and their use in economics). The advantage of this modelling approach is that it is simple and computationally tractable, while capturing differential linking across groups that is essential to our analysis.

priors **correlated priors**: holding fixed average precision ($\frac{s_1+s_2}{2}$), the larger $C = |s_1 - s_2|$, the stronger the correlation of priors.

The social learning unfolds as follows. At the end of period $t = 0$, each agent's action is simply their private signal, $a_i^0 = \omega_i$. In each period thereafter, agents play a *myopic coordination game* with their network neighbors. In this game, agents get utility when repeating their own past action, but also for coordinating with the actions of other, which depends the identity of observed individuals. Formally, for $t \geq 1$, agents' utilities are:

$$u_i(a_i^t) = -\frac{1}{d_i}\left(\underbrace{|a_i^t - a_i^{t-1}|}_{\text{Own action}} + \underbrace{(1-b_{si})}_{\text{Same group weight}} \underbrace{\sum_{j:i\neq j,k_i=k_j} |a_i^t - a_j^{t-1}|}_{\text{Same group actions}} + \underbrace{(1-b_{di})}_{\text{Diff. group weight}} \underbrace{\sum_{j:i\neq j,k_i\neq k_j} |a_i^t - a_j^{t-1}|}_{\text{Diff. group actions}}\right) \quad (4)$$

where $d_i = \sum_j G_{ij}$. The bias parameters have an natural interpretation, for example, if $b_{si} = 0.6$, then i follows same-group actions 40% as much as they follow their own action. In the model, we define **bias** as $b = 1 - \frac{1-b_d}{1-b_s}$, which measures how much more individuals discount information from different groups versus information from the their own group. We restrict attention to cases in which $b_s, b_d \in [0, 1]$ and $b_s \leq b_d$, such that $b \in [0, 1]$.[16]

Before characterizing the best-reply dynamics to this game, define an $n \times n$ updating matrix $\mathbf{T}(\mathbf{G})$ with characteristic entry

$$T_{ij} = \frac{G_{ij}}{\sum_i G_{ij}} \quad (5)$$

i.e. $\mathbf{T}$ is a row-normalized version of the adjacency matrix $\mathbf{G}$ that represents the network. Then, the social learning rule for $t \geq 1$ that arises as a best-reply to the coordination game with utilities in Equation 4 is:

$$\mathbf{a}^t = sgn\left[\mathbf{T}(\mathbf{G})\mathbf{a}^{t-1}\right] \quad (6)$$

where *sgn* is the signum function that is applied element-wise; this returns 1 if the element is positive, -1 if

[16]This means that on average players don't get positive utility from mismatching action, and that on average they don't follow others' information more than their own (note that individual agents may still have $b_{si}, b_{di} > 1$ or $b_{si}, b_{di} < 0$)

the element is negative, and the previous-period action if the element is zero. That is, for a typical agent i this updating rule implies a weighted average of previous-period actions $a_i^t = sgn\left[\sum_j T_{ij}(\mathbf{G})a_j^{t-1}\right]$. This social learning rule corresponds to DeGroot updating with bias.[17] Importantly, the weighted average is proportional to the weights of the links in the network $\mathbf{G}$; this means that the larger the bias parameter b_s, b_d, the more agents discount the actions of individuals from the same and different group respectively.

3.2 Hypotheses

To formalize testable hypotheses, we are interested in how the effects of integration $I = p_d/p_s$ on social learning is shaped by bias $b = 1 - \frac{1-b_d}{1-b_s}$ and the correlation of priors $C = |s_1 - s_2|$. While the model is simple and intuitive, it is analytically intractable for even moderately sized networks; one has to account for all possible networks that can form under the parameters and all possible schedules of signals, calculate the probabilities of each of these events and the learning outcomes associated with them, and then integrate over all such possibilities. In order to go beyond these analytical difficulties, we run simulations to formalize testable hypotheses.

The simulation algorithm is detailed in the Algorithm 1 box below. The inputs for the algorithm are population size n, probability of same-type and different-type links p_s, p_d, precision of priors for each group s_1, s_2, the mean bias towards the same and the different group b_s, b_d and their SDs σ_s, σ_d, and number of updating rounds in the learning stage L. For each set of these parameters, we iterate the model H times. The simulations essentially run the stochastic model: first, we generate random networks, then, we seed random priors, and finally play out learning using the updating rule implied by the coordination game above. The outcome of interest is the fraction of agents that learn the true state of the world, and how this changes as we increase integration.

In the main results we set parameters as follows. For the bias parameters, we set no same-group bias

[17] The DeGroot model has a long history – see French Jr., 1956, Harary, 1959, DeGroot, 1974 for the earliest formulations of the rule, and Golub and Sadler, 2016 and Jackson, 2008 for great reviews of this model and its relation to economics. We favor this model for two main reasons. First, there is growing empirical evidence that Bayesian learning is not a good predictor of real life behavior, and that the DeGroot model may do a better job (Massey and Wu, 2005, Weizsäcker, 2010, Chandrasekhar et al., 2020). Second, the parameters in the model have a natural interpretation, and we can use it to estimate the bias parameters b_s, b_d. The model is also computationally tractable in large real-life networks, and facilitates simulations and counterfactual excesses that are of interest to our study.

Algorithm 1 Theoretical Simulations of Social Learning

Inputs: $\{n, p_s, p_d, b_s, b_d, \sigma_s, \sigma_d, s_1, s_2, L, H\}$
for $h \in \{1, 2, ..., H\}$ **do**
    Generate a random network $\mathbf{G}(n, b_s, b_d, \sigma_s, \sigma_d, p_s, p_d)$ as described in Equation 1 and Equation 2
    Generate random vector of initial beliefs $\omega(s_1, s_2)$ as described in Equation 3
    Generate an updating matrix $\mathbf{T}(\mathbf{G}(n, b_s, b_d, \sigma_s, \sigma_d, p_s, p_d)$ as described in Equation 5
    Assign $\mathbf{a}^0 = \omega$, and calculate final learning vector $\mathbf{a}^L$ using Equation 6
    Calculate network-level learning for iteration h as $a_h = \frac{\mathbf{1}'\mathbf{a}^L}{n}$
    Store a_h
end for
Calculate average learning over iterations $a = \frac{\sum_{i \in \{1,2,...H\}} a_i}{H}$
Store the mapping $(n, p_s, p_d, b, \sigma, s_1, s_2, L) \rightarrow a$

$b_s = 0$, and simulate results for a grid of different-group biases $b = b_d \in \{0, \frac{1}{3}, \frac{2}{3}, 1\}$. We also set $\sigma_s = \sigma_d = 0.35$.[18] We vary integration by fixing expected number of links d, and vary p_s, p_d as follows: $(p_s, p_d) \in \left\{ \left(\frac{d}{(I+1)n}, \frac{Id}{(I+1)n} \right) : I \in \{0, 0.1, 0.2, 0.3, ..., 0.9, 1\} \right\}$, where I denotes integration. For the signal distribution, we set average precision as $\frac{2}{3}$, and consider three levels of correlation of priors: $s = (s_1, s_2) \in \{(0.66, 0.66), (0.76, 0.56), (0.86, 0.46), (1, 0.33)\}$.[19] In our main results, we set $n = 100, d = 5$. We chose the average number of links d in relation to the average precision in order to avoid corner cases. We set $H = 1000$, $L = 10$. In general, these parameters should be as large as possible for the law of large numbers to kick in (in the case of H) or learning to stabilize (in the case of L). We find that these choices yield stable estimates.[20] The results seem robust to alternative parameter specifications (see Appendix A.1).

The simulation results are presented in Figure 3. On the x-axis we plot integration I, and on the y-axis we plot the average learning a as a fraction of individuals who correctly learn the true state of the world. To more formally discuss the results, let $L()$ denote learning, where the brackets include different parameters of the learning process. Fix any two points on the x-axis $I_L < I_H$, and define the benefits of integration as $B = L(I_H) - L(I_L)$. We are interested in how B changes for different bias and correlation of priors.

How does bias matter? In Panel A, we present how learning changes over bias. We can notice two patterns. First, bias weakly decreases learning. As bias increases, each line shifts downwards. This is not surprising,

[18] Varying these makes the estimates for each set of parameters more or less precise, and it does not ordinal rankings between outcomes across different parameters.

[19] In general, the larger this is the more learning overall, and vice versa. For large enough or small enough precision, we create a corner case in which there is perfect or no learning. In these cases, segregation, bias, and correlated priors do not matter and learning is the same for any parameters.

[20] Though note that convergence is not guaranteed with discrete learning. Under some realized networks, there could be some local oscillations, i.e. most agents converge to a belief, while few alternate their guesses.

since higher bias means less information is incorporated. We can also notice that bias decreases the benefits of integration – B goes down as bias goes up. We can notice that in an environment when integration leads to more learning, the gains in learning are reduced when the bias is larger. Conversely, in an environment where integration decreases learning, bias makes that decrease larger. The intuition here is simple too – if agents incorporate less information from cross-group individuals, integration will have a more muted effect on learning, either diminishing positive changes, or magnifying negative changes.

How do correlated priors matter? In Panel B, we present how learning changes over correlation of priors. We can notice that correlated priors decrease overall learning, the more different the initial precision of the signal across the two groups, the smaller the overall learning. We can see this on Panel B, as the lines associated with larger correlation of priors are lower. The intuition here is that as we make the signals more dispersed across groups, the gain in precision in the better-informed group does not make up for the loss of precision in the worse-informed group.[21] We also see that correlated priors increase the benefits of integration. As the correlation of priors increases, B increases as well. The intuition here is that when one group has better information than the other, integrating the groups can improve overall learning since it allows the correct information to flow, even if bias is high, avoiding echo chambers.

Overall, the model simulations imply the following testable hypotheses:

Hypothesis 1.

1.1. Bias weakly decreases learning.

1.2. Bias weakly increases the costs of integration on learning.

Hypothesis 2.

2.1. Correlated priors weakly decrease learning.

2.2. Correlated priors weakly increase the benefits of integration on learning.

[21] This conclusions however hinges on the assumption that the signal is correct on average (i.e. $\frac{s_1+s_2}{2} > \frac{1}{2}$), a common starting point in these models in the literature which we follow here. If the signal is correct less than half of the time on average, then correlated priors *decrease* the benefit of integration. Intuitively, if the signal is poor on average, concentrating it in one tight knit group guarantees higher spread compared to spreading it across two such groups, since networked learning exhibits threshold behavior. Appendix A.1 shows this case.

In Appendix A.2, we present an extension of the model in Golub and Jackson, 2012, a seminal paper that studies how integration improves the speed of learning. While our modelling strategy mirrors theirs, there are two key differences. First, they consider asymptotic learning, such that society grows infinite, and use continuous updating with an uncountable set of actions. These two features allow tractable theoretical treatment of this question. Second, in their their model there is no bias ($b_s = b_d = 0$), while our model introduces the idea that groups, besides simply being segregated in the social network, could also differentially incorporate information from others. In our extended model where $b_s, b_d \geq 0$, we show that the speed of learning depends on both integration *and* bias, introducing some nuance in the interpretation of this finding.

4 Study design and data

4.1 Study design

Implementation

Participation in the study was open to all students during recruitment. To recruit students, local enumerators fluent in the students' native language went to their classrooms and presented the opportunity to participate in the study. The enumerators then helped students fill an assent form, after which their parent received an automatic email with a link to fill in parental consent online. To avoid experimenter demand effects and reduce self-selection into the study, we framed the study around friendship and social learning.[22] The lab game was designed for groups of 6 students – 3 students from 2 distinct ethnicities. For example, in some sessions, a group was composed of 3 Macedonians and 3 Albanians, in others 3 Albanians and 3 Turks, etc. The summary statistics for the participants are presented in Table 1.

Each gaming session was led by a pair of two facilitators. The facilitators were school teachers who were hired and trained according to a protocol. The sessions were held with students in the same room playing simultaneously and the students could use either a computer or a smartphone, both of which were compatible with our platform. Prior to each session, we randomly chose players to compose the group. To ensure

[22] The consent language stated: *"We are conducting a study to understand how students make friends and learn about educational opportunities. Participation includes answering questions and participating in games. As part of the study, you may have the opportunity to participate in free tutoring/school enrichment in English."*

compliance on sessions composition, we asked participants before the game to report their birth month and year, and matched it to the birth month and year reported at registration. We find that 96.03% of the players have a perfect match, suggesting a high compliance with the random composition of sessions. In case of a participant no-show, the facilitators were provided with two randomly chosen reserve players, one from each ethnicity. In case the reserve players were also unavailable, the session was rescheduled.

We built a custom platform for the game.[23] This platform allowed us to set-up gaming sessions with a pre-determined composition of players and facilitated the randomization as described later in this subsection.

Learning game

In the game *Red or Blue?*, participants are told that there are two bags, a red bag (with 4 red balls and 2 blue balls) and a blue bag (with 4 blue balls and 2 red balls). The computer picks one bag at random, and the goal of the game is to correctly guess the color of the chosen bag. To help players guess, the computer randomly distributes one ball from the bag to each of the 6 students. Thus, 4 students get a correct signal and 2 students get an incorrect signal about the true state of the world. In the first round the students can see their ball and guess the color of the bag. In subsequent rounds, each player sees their own guess from the previous round, along with the guesses that two or three other players made in the previous round. After observing this, players make a guess in the current round. The students see the full names of the players they observe, along with the color they truly selected, shutting down the possibility of cheating or strategic communication.

We took a few steps to maximize comprehension of the game. At the beginning of each session, the students watched an animated video in their native language that explains the rules of the game; the English language version of the video is available on this Youtube link.[24] Throughout the session, we showed participants a board-game-style card that reiterated the main goals and rules of the game. Finally, we allowed the students to play two practice games to become familiarized with the game mechanics. We discard data from the practice games in all our analyses.

Each group plays 6 games on average[25], and each game repeats for 3 to 5 rounds, stopping randomly to avoid

[23] The python-based platform is available on GitHub here: https://github.com/prijatelilab/PrijateliTree

[24] https://www.youtube.com/watch?v=h4S1J68MifE

[25] Additional 6 on average were played in an anonymous treatment during which we did not display students' names. We use this

end-of-experiment effects. The students are incentivized to guess the correct color: if they pick the correct color in the last round of the game they get a cash reward, otherwise they get nothing. Every 250 points are worth 50 MKD (0.9 USD). Points sum over the games, and students win rewards for all collected points. On average, each game win paid 27 MKD (or about 0.45 USD). To put this in context, with 3 wins a student has a little more money than is needed for a coffee in a typical coffee shop. During the sessions, participants are only allowed to watch their screens and verbal communication about the game is strictly forbidden. If students break this rule, the facilitator is responsible to disqualify them, though this did not happen in practice and teachers report little rule-breaking.[26]

Experimental manipulation

The lab game incorporates a randomization design that varies the three parameters in the model, and helps us test the hypotheses. First, we randomize the network structure on which the 6 students can observe each other. Figure 1 shows the *Segregated* and *Integrated* treatments. Students are represented as nodes, and distinct ethnicities are denoted by different colors/shapes. Each link denotes a bidirectional observation during the game. The networks capture the network formation model well. The segregated network has $p_s = 1, p_d = \frac{1}{9}$, and the integrated network has $p_s = \frac{2}{3}, p_d = \frac{1}{3}$. Importantly, we fix the number of links for each network position and each individual, while varying the identity of the observed players. For example, Player 2 in Figure 1 can observe 2 players in both networks; in the segregated network they can see 2 players of their own ethnicity, while in the integrated network they see one player from each ethnicity. Therefore, the amount of information available to each player is the same in expectation, but the identity of the observed individuals changes. We administer this manipulation at the game level, and each arm is selected with 50% likelihood. To ensure comparability we also randomize the network position of each player across games, though hold it fixed within game and across rounds.

It is important to note that under no bias, $b_s = b_d = 0$, where each observed information is given equal weight, the model predicts that the integrated network should lead to more learning than the segregated network. Under the model, the average learning in the integrated network over all possible initial signals is 84.4%, while the average learning in the segregated network is 80%. This is because the segregated

arm as a placebo check in our results section. Each group played exactly 12 games, and each game had 50% chance of showing the players' names

[26] 51% say there was no talking at all, and 93% of facilitators rank the amount of student talking to be 3 or less on a 1-6 scale. When asked whether the talking was explicitly about the game and broke the rules, and received only 2 reported instances.

network is more prone to create echo chambers: if one ethnicity gets 2 incorrect and 1 correct signals in the network, the model predicts that each cluster averages to a different action and perpetually gets stuck there. This happens with lower likelihood in the integrated network since information can flow more easily among players. Therefore, the experiment induces an overall environment in which integration is optimal if the students do not have any bias. This is consistent with previous findings in the literature on the benefits of integration by avoiding echo chambers (Golub and Jackson, 2012, Chandrasekhar et al., 2020).[27]

Second, in each game we randomize the correlation of prior signals. Since 4 out of 6 players get a correct signal, how these signals are distributed across ethnicities varies the correlation of priors. The *uncorrelated priors* arm are the games in which both ethnicities have the same information, which in the model maps to $s_1 = s_2 = \frac{2}{3}$, and each ethnicity gets 2 correct and 1 incorrect signals. In the *correlated priors* arm, one ethnicity will receive 3 correct signals, while the other ethnicity will have 2 incorrect signals and 1 correct; in the model this maps to $s_1 = 1, s_2 = \frac{1}{3}$ (or $s_1 = \frac{1}{3}, s_2 = 1$). Under correlated priors, one ethnicity has worse information and can benefit from observing the other ethnicity that has better information. Initial signals are randomized at the game level. Since we distribute signals by random shuffling, there are 6 possible schedules of signals that are in the correlated priors arm, and 9 possible schedules that are in the uncorrelated priors arm. Therefore, the first treatment is administered with 40% probability and the second with 60% probability.

Lastly, we induce random manipulation of bias by randomly composing the groups and inducing a variation in the "types" of players. Before the games, we administered a network survey in which students report friendships to other students in their school. Only 45% of participants report a cross-ethnic friendship. If a student reports at least one cross-ethnic friend, we designate their type as "low-bias", and if they report no cross-ethnic friends, they are a "high-bias" type. Since groups are randomly formed, some groups have more players with cross-ethnic friends than other groups by chance. We split the groups by the median of this group-level bias variable, thus defining *High Bias* groups, in which 0-3 players have a cross-ethnic friend (about 60% of groups) and *Low Bias* groups, in which 4-6 students have a cross ethnic friend (about 40% of groups).

This manipulation of bias comes with two limitations. First, while the composition of players is random, whether a student has cross-ethnic friends can be correlated with other characteristics. For example, if hav-

[27] We show results for both overall learning, and for learning relative to the model-predicted no-bias learning, therefore "normalizing" for the network structure.

ing more cross-ethnic friendships is also associated with higher general trust, it may be trust rather than lower bias that matters. In the results below, we unpack what other observables are associated with having a cross-ethnic friend, and rule out some competing explanations. Second, the "name treatment" is a bundled treatment. During the game, we show the names of other players along with their guesses, and rely on the strong association between names and ethnicity to detect differential updating by ethnicity. While names strongly connote ethnicity in our context, they can also signal other aspects of one's identity, such as gender. Later in the paper we estimate the learning bias toward the same-group and different-group individuals, so it is important to keep in mind that we cannot interpret this as ethnicity bias per se. We follow philosopher Miranda Fricker to interpret this as a less narrowly defined *identity prejudice*, where identity in our case is perceived via the name of the individual (Fricker, 2007). In Sections 5 and 6 we take several steps to alleviate concerns related to these limitations in the bias manipulation.

Our design has a couple of core advantages. First, it addresses the empirical problems associated with studying our question. By randomizing the network structure and the initial signals, we deal with the correlated effects problem, since we eliminate common characteristics driving initial actions as well as endogenous selection of peers that can confound estimates in observational studies. Additionally, we eliminate worries about asymmetric preferences. Students in our sample show similar preferences for the colors: the average percentages of pressing Red-Blue are 54%-46% for Albanians, 54.4%-45.6% for Macedonians, and 54.75%-45.25% for Turks. The second advantage of the game is that it mirrors our model and explicitly randomizes the model parameters. This allows for a clear mapping between model-implied hypotheses and empirical tests. Furthermore, we can use the model to estimate the bias parameters b_s, b_d from the game play data. This allows us to combine these parameter estimates with data on real-world networks, and simulate policy counterfactuals that are relevant for local policymakers.

4.2 Data

We use several administrative and original data sources in our analyses. Our main data source is the online game platform which we built and designed with custom features to fit our needs. This platform exports complete data about all features and interactions during each game: an ID to link participants from sessions to other data, data about the game including all the random treatments (true color, signals, treatment conditions,

network positions, etc.) and each participant's action in each round. The sessions began in April 2024 and were finalized in December 2024.

We administered an original baseline survey before the games (relevant baseline characteristics of the sample are shown in Table 1).[28] Beyond a standard demographic module, the baseline survey included several other modules. First, a module elicited inter-group preferences. We asked students to play two dictator games against an anonymous peers with same mother tongue (in-group) and another two games against an anonymous peer with a different mother tongue (out-group).[29] We measure *Dictator Game bias* as the difference in cash distributed to the peers of same versus different mother tongue, averaged over questions. We also administered an Inclusion of Others in Self survey question following Aron et al., 1992, in which students report perceived proximity to each local ethnic group. We define *Proximity to other local ethnicities* as the average report for all local ethnicities, excluding the student's own ethnicity. We also administered a Stereotype Content Model module (Cuddy et al., 2008), which measures how students rank ethnic groups on a warmth and competence scale (these stereotypes are plotted in Figure 2).

Second, the baseline survey included language module in which students list and rank, in order of fluency, Macedonian, Albanian, Turkish, and English. For each student, we calculate *Fluency in other local languages* as the average reported fluency of all local languages, excluding the student's native language. Third, a module on attitudes, including life satisfaction and *Generalized trust*, in which students report how trustworthy people are in general in their society. Fourth, we included a parent module, in which students reported their parents' education level.

Finally, our survey included a social networks module. In a random network census, we randomly draw 12 students from the school and ask whether students would say hi to them. In a named network census, we ask students to list up to 6 peers in four types of ties: close friends (clique), peers they go to for school help, peers they go to to get information, and cross-ethnic friends. We aggregate all four questions to construct school-cohort level networks, and present network statistics in Table 2. In the penultimate chapter, we use these real-life networks to conduct counterfactual policy exercises.

We also administered two short Qualtrics surveys during the gaming sessions. One is administered to students

[28] The baseline survey happened before the games for 99.6% of the participants with only one exception to the protocol.
[29] The questions also varied whether the peers are from the same or a different school.

directly on the gaming platform. This survey includes the animated explanation of the game, simple tips for playing, and asks participants to confirm their birthdays. The second survey is filled by the session facilitators. After each session, the facilitators submit a survey in which they indicate whether the students broke communication rules during the session, the degree to which each student seemed to understand the game, and the degree to which each student seemed to enjoy the play.

Finally, we have school data on students' GPA, classroom assignment, subjects taken, and class absences.

5 Testing the hypotheses

How much learning is there?

We find that on average, games end with 4.54 of the 6 players guessing correctly. Since the game starts with 4 correct and 2 incorrect signals, this means that the fraction of agents who correctly guess goes from 66.66% to 75.66%, an increase of 9% in learning. However, there is considerable heterogeneity across games. Figure 4 plots the histogram of the number of agents who correctly guess in the last round. The modal outcome of the games is perfect learning: all six players guess correctly in 45.3% of the games! Yet, almost 28% of all games end in mislearning: fewer than 4 players guess the color correctly. This outcome is worse than if players guessed in isolation and followed only their private signal. This heterogeneity in outcomes motivates our analyses: when are agents more likely to end in the world of perfect learning, rather than in the world of mislearning?

Testing Hypothesis 1: Bias, Integration, and Learning

Recall that Hypothesis 1 stated that bias decreases learning and increases the costs of integration. Specifically, we want to test:

$$L(HighBias) \leq L(LowBias) \tag{7}$$

$$L(HighBias, Int.) - L(HighBias, Seg.) \leq L(LowBias, Int.) - L(LowBias, Seg.) \tag{8}$$

where $L()$ represents average learning under the condition within the brackets. To test these two implications, we focus on network-level learning, measured as the fraction of players who guess correctly in the last round of each game. We compare groups in which 3 or fewer students have at least one friend from a different ethnicity (*HighBias*) with groups in which 4 or more students have at least one friend from a different ethnicity (*LowBias*). This represents a median split with respect to our bias proxy, which we manipulate at the session-level by randomly composing the group of 6 players.

Figure 5 plots the raw means associated with the inequalities in Equations 7 and 8. In Panel A, we plot the means by *Low Bias* and *High Bias* to test the implication in Equation 7. Panel B gets at the claim in Equation 8 by splitting the means additionally for the *Integrated* and *Segregated* treatments.

Figure 5 suggests that the means are consistent with the model predictions. To formally test the hypothesis, we run the following regressions:

$$Learning_{ig} = \beta_0 + \beta_H High_i + X_i + \sigma_i + \epsilon_{ig} \tag{9}$$

$$Learning_{ig} = \beta_0 + \beta_{HS} HighSeg_{ig} + \beta_{LI} LowInt_{ig} + \beta_{LS} LowSeg_{ig} + X_i + \sigma_i + \epsilon_{ig} \tag{10}$$

where the i denotes session and g game; $High$ (Low) are dummies that take the value of 1 when the group in the session has below-median (above-median) cross-ethnic friends; and Seg (Int) are dummies that take the value of 1 when the game is played on a segregated (integrated) network. The regressions also add fixed effects X_i for school-by-grade randomization strata, and ethnic pairing (three indicators for Macedonian-Albanian, Albanian-Turkish, and Turkish-Macedonian). We cluster standard errors at the session-level, which is the highest level of randomization in our design. The coefficients of this regression allow us to formally test Hypotheses 7 and 8; Equation 9 formalizes the results presented in Figure 5 Panel A, and Equation 10 formalizes the results presented in Figure 5 Panel B.

Table 3 shows the results of these regressions. We take a standard hypothesis testing approach. Specifically we test:

$$H_0 : \ \theta > 0, H_A : \ \theta \leq 0 \tag{11}$$

where $\theta = \beta_H$ in Equation 9 (as implied by Hypothesis 7), while $\theta = \beta_{LS} - \beta_{LI} - \beta_{HS}$ in Equation 10

(as implied by Hypothesis 8). Below each regression in Table 3, we present the estimated θ and the $p-$value associated with the one sided hypothesis test that the observed data is generated under the null. As predicted by the model, all the estimated θ have a negative sign. Moreover, we can reject the null that the observed data is generated under the null consistently at the 10% level. We conclude that the observed behavior in the experiment is consistent with the theory: higher bias reduces learning, and increases the costs of integration.

It is important to note that we cannot conclude that making a cross-ethnic friend makes one learn better in integrated networks. Rather, our interpretation is that there are heterogeneous effects of integration by type. We think of each individual as having a *type* defined by their bias, as proxied by cross-ethnic friendships. We contend that the optimality of integration depends on the existing bias of the student. Still, it may be that other characteristics interact with the integration treatment. For example, if students with cross-ethnic friends also have higher trust, it may be that we are capturing a trust channel in our results. In Table 4, we correlate our proxy of type (cross-ethnic friends) with other observed characteristics. We find that having a cross-ethnic friend is related to lower in-group bias in a Dictator game, higher fluency of non-native local languages, and perceiving other local ethnicities as closer on the Inclusions of Other in self metric. These are all characteristics that we expect from someone with prior interaction with other ethnicities, consistent with the type we are trying to capture.

The table also suggests that we can reject at least two competing explanations for our results. First, individuals with cross-ethnic friends are not more trusting generally; on the contrary, they show lower levels of generalized trust, though this relationship is not significant. Therefore, the optimality of integration over segregation for the low bias types cannot be explained by generalized trust. Second, we see that having a cross-ethnic friend is associated with having less educated parents, lower educational ambitions, and lower GPA. To the extent that these capture cognitive abilities, we can reject the alternative explanation that cognitive skills lead to better learning in integrated networks.

Testing Hypothesis 2: Correlated Priors, Integration, and Learning

We take an analogous approach to testing Hypothesis 2. Hypothesis 2 suggested that correlated priors decrease learning, but increase the benefit of integration. Specifically, we want to test that:

$$L(Corr) \leq L(Uncorr) \tag{12}$$

$$L(Uncorr, Int.) - L(Uncorr, Seg.) \leq L(Corr, Int.) - L(Corr, Seg.) \tag{13}$$

Figure 6 plots the means associated with the two parts of Hypothesis 2 stated above – splitting learning by the *Correlated* and *Uncorrelated* treatment in Panel A, and then additionally by the *Integrated* vs. *Segregated* treatment in Panel B. Additionally, we put these raw means in analogous regression forms:

$$Learning_{ig} = \beta_0 + \beta_C Corr_{ig} + X_i + \sigma_i + \epsilon_{ig} \tag{14}$$

$$Learning_{ig} = \beta_0 + \beta_{CS} CorrSeg_{ig} + \beta_{CI} CorrInt_{ig} + \beta_{UI} UncorrInt_{ig} + X_i + \sigma_i + \epsilon_{ig} \tag{15}$$

where $Corr$ ($Uncorr$) are dummy variables that equal 1 when the game has the *Correlated* (*Uncorrelated*) priors treatment, and Int (Seg) are dummies that are 1 when the game has the random *Integration* (*Segregation*) treatment. We include the same fixed effects in X_i and again cluster at the session level. Table 5 shows the output of these regressions. We take the same approach of testing the one-sided null hypothesis as in Equation 11. In this case, we want to set $\theta = \beta_C$ for the regression in Equation 14, while $\theta = \beta_{UI} + \beta_{CS} - \beta_{CI}$ for the regression in Equation 15.

Below each model in Table 5 we present the estimated θ and the p-value associated with the one sided hypothesis test that the observed data is produced under the null. Again, as predicted by the theory, all estimated θ are negative. While we do not find statistically significant evidence for the first part of the claim that correlated priors reduce learning, we do find statistically significant evidence at the 10% level for the claim that correlated priors increase the benefits of integration.

Echo chambers

One advantage of our experiment is that we can observe emergent network phenomena. Since groups in the real world often diverge in beliefs and behaviors, echo chambers are of particular importance to our question. In our study we define echo chambers as the event when the ethnic groups diverge in their guesses; that is, at the end of the game all three students of one ethnicity choose red (blue) while all three students of the other ethnicity choose blue (red). This happens relatively infrequently, with only 9.23% of all games ending with echo chambers (compared with 45.3% of all games that end with perfect learning).

We expect that high bias leads to more echo chambers, and that segregation leads to more echo chambers when bias is high. Figure 7 shows the raw means of games that end in echo chambers by network structure and bias. Analogous regressions to the ones in Equations 9 and 10 are shown in Table 6. In this case we expect that $\theta = \beta_H \geq 0$ and that $\theta = \beta_{HS} - \beta_{LS} + \beta_{LI} \geq 0$. The θ at the bottom of each model is the p-value that the experimental data was drawn from a distribution under the null hypothesis that $\theta < 0$. While we see that θ is positive as expected across specification, we cannot reject the null at any conventional significance levels.

We also expect that correlated priors increase echo chambers, since they already set up the groups with unequal information by definition; for similar reasons we also expect that the segregation increases echo chambers more under correlated priors. Again, we show analogous figures of the echo chambers across treatment conditions in Figure 8. We run analogous regressions to those in Equations 14 and 15, predicting that $\theta = \beta_C \geq 0$ and $\theta = \beta_{CS} - \beta_{CI} + \beta_{UI} \geq 0$. We see similar results as above: the results are directionally consistent with our expectations (positive θ), but we cannot reject the null that θ is negative at any conventional significance levels.

Robustness checks

Our results are robust to multiple robustness checks, which are shown in Appendix B. First, in some games we administered anonymous play, in which we did not reveal students' names during the play, but instead used code names like "Player 1". We use these games as placebo checks – since students cannot distinguish

between players, we expect that the results are inconsistent with the hypotheses about the costs and benefits of integration. As predicted, in Appendix B.1 we show that the results generated under anonymous play are inconsistent with the theoretical hypotheses, suggesting that the observation of names is driving the observed results.

In Appendix B.2, we show that the results are similar if instead of the last round, we use only the first round, or all rounds of each game. In Appendix B.3 we show robustness to alternative bias proxies. Instead of using reports of cross-ethnic friends, we proxy for bias using the Dictator Games, Inclusion of Others in Self metric, or reported fluency of non-native local languages to proxy for bias (these variables are described in detail in Section 4). The IOS metric shows inconclusive evidence for our hypotheses, but the language fluency and dictator games show very similar results, and the Dictator Games show even stronger evidence for the hypothesis that bias increases the costs of integration. Moreover, we test robustness to the results by tercile split on the bias proxy (as opposed to median). We run regressions where we compare the second to first tercile, and the third to first tercile of groups according to their group-level bias. The results remain unchanged, bias still predicts lower learning and smaller benefits of integration on learning. Finally, in Appendix B.4, we repeat our results with the normalized learning outcome. This is calculated as the difference between the actual learning and the learning predicted under the model in the given network under no bias. In essence, this calculates deviations from the no-bias learning in segregation (0.80) and integration (0.84), thereby normalizing learning with respect to the given network structure. The results are qualitatively the same.

Overall, our formal tests and robustness checks show that the data behaves consistently with the hypotheses. These empirical findings provide evidence for the hypotheses and imply a nuanced effect of integration. First, the benefits of integration increase when groups have correlated priors, such that one group has to learn from the other. Second, the costs of integration increase as player bias increases. The results suggest that in practice, it is this interaction of bias and prior information that may decide whether the overall effects of integration and positive or negative.

6 Model fitting and counterfactuals

6.1 How biased are the students?

In this section, we estimate the learning bias students exhibit in the lab game. In the model presented in Section 3, the utility of each individual i depends on coordinating actions with same- and different-group individuals. The utility was driven by bias parameters $b_i = (b_{si}, b_{di})$, such that the larger the bias, the smaller the disutility in failing to coordinate with players of the respective types. Rewrite the utility of an individual i choosing action a at time t from Equation 4 as:

$$u_i^t(a_i^t, b_i) = -\frac{1}{d_i}\left(\mathbb{1}[a_i^t \neq a_i^{t-1}] + (1 - b_{si})s_{si}^t + (1 - b_{di})s_{di}^t\right) \tag{16}$$

where $s_{si}^t = \sum_{j:k_j=k_i} \mathbb{1}[a_i^t \neq a_j^{t-1}]$ is the number of same-group individuals that i observes that chose the opposite action than a_i^t in the previous period, and analogously for $s_{di}^t = \sum_{j:k_j\neq k_i} \mathbb{1}[a_i^t \neq a_j^{t-1}]$. Recall that $d_i = \sum_j G_{ij}$. Since we normalize own-bias to zero, the bias parameters have a natural interpretation. For example, if $b_{si} = 0.4$ then i values the information from the actions of same-group agents $1 - b_{si} = 0.6$ times as much as they value their own; in other words, they incorporate 60% of the information of same-group individuals.

For each agent, we assume that the individual bias parameters are drawn from normal distributions: $b_{si} \sim N(b_s, \sigma_s)$, $b_{di} \sim N(b_d, \sigma_d)$, or more succinctly as $b_i \sim f(b_i|\theta)$. The goal is to estimate the parameter set $\theta = (b_s, \sigma_s, b_d, \sigma_d)$ that governs the individual biases. We take a mixed logit approach to estimate these biases, in which we model the choice selection as probabilities that depend on the utility function above and extreme value errors (Train, 2009). Let $x_i^t = (a_i^{t-1}, s_{si}^t, s_{sd}^t)$ describe the choice environment that i is facing at time t, and define x_i as the choice environment i faces across all time periods. Let y_i be the sequences of actions made by agent i, under choice environments x_i. Since the mixed logit model assumes an extreme value distribution of the errors, the probability of the agent taking the sequence of actions y_i is:

$$\mathbb{P}(y_i|\theta, x_i) = \int L_{y_i}(b_i) f(b_i|\theta) \tag{17}$$

where $f(b_i|\theta)$ is the bivariate normal distribution of bias parameters, and L_{y_i} is the standard logit probability

that i chooses a sequence of actions y_i under choice environment x_i and bias parameters b_i:

$$L_{y_i}(b_i) = \prod_t L_{a_i^t}(b_i) \tag{18}$$

$$L_{a_i^t}(b_i) = \frac{exp(u_i(a_i^t, b_i))}{exp(u_i(a_i^t, b_i)) + exp(u_i(\neg a_i^t, b_i))} \tag{19}$$

Estimating $f(b_i|\theta)$

The first step is to estimate the parameter set θ that governs the distribution of biases in the population as a whole. We estimate these parameters through a Maximum Simulated Likelihood Estimator (MSLE). More specifically, for a given θ we estimate the probability of a sequence of actions y_i under choice environment x_i as:

$$\hat{\mathbb{P}}(y_i|\theta, x_i) = \frac{1}{R}\sum_r L_{y_i}(b_i^r) \tag{20}$$

where each b_i^r represents one draw from $f(b_i|\theta)$. The simulated log-likelihood given parameters θ is:

$$SLL(\theta) = \sum_{i \in N} ln(\hat{\mathbb{P}}(y_i|\theta, x_i)) \tag{21}$$

and the MSL estimator finds the parameters that maximize this object, or:

$$\hat{\theta} = argmax_\theta SLL(\theta) \tag{22}$$

We implement the estimation in Pyhton, using the built-in scipy optimization program.[30] In our main results we set $R = 100$, and bound the parameters as $b_s, b_d \in [-1, 2]$ and $\sigma_s, \sigma_d \in [0, 1]$.

The results are shown in Table 8. The average same-group bias is 0.5, and the average different-group bias is 0.79. This means that on average players weigh information of co-ethnics 0.5 times as much as they weigh they own information; in contrast, they weigh the information of cross-ethnics only 0.21 times as much as they weigh their own information. In short, players are biased against everyone, but they are discount cross-ethnic

[30] We apply the limited memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, i.e. the L-BFGS-B option in Python.

information more than twice as much as they discount co-ethnic information.

It is useful to put these numbers in context. That people weigh the information of others less than they weigh their own information is a known finding. In fact both Mobius et al. (2015) and Conlon et al. (2022) find that people weigh others' information about 0.4 times as much as they weigh their own information. This is not very different from our estimate which is 0.5, despite the fact that all of our studies have different subject populations and experimental designs. Conlon et al. (2022) interpret this discounting as an ownership effect, and they "rule out distrust, confusion, errors in probabilistic thinking, up-front inattention and imperfect recall as channels" (p. 1). However, our results suggest a limitation of this interpretation. While the same-group bias can perhaps be explained by an ownership effect channel, we estimate a much smaller learning from cross-ethnics. The gap between same-group and different-group bias cannot be explained simply by an "ownership effect". Instead, it suggests identity as an important channel.

Our interpretation follows what philosopher Miranda Fricker labels as "identity prejudice", which is described as a prejudice against people because of who they are (Fricker, 2007). Identity prejudice gives rise to testimonial injustice, which "occurs when [identity] prejudice causes a hearer to give a deflated level of credibility to a speaker's word".[31] Identity prejudice imposes costs on both sides of the exchange: it undermines the speaker's credibility and agency as an information provider, while simultaneously reducing the hearer's ability to update their beliefs or acquire knowledge—resulting in a breakdown of efficient information transmission. We favor this interpretation because it is both specific enough to explain the effect we find, and broad enough not to imply anything about specific channels beyond the scope of our study. One consequence of this interpretation is that it begs more research on how identity matters for social learning.

Estimating individual-level bias parameters

Now we estimate the bias parameters at the individual level, an exercise which serves two goals. First, we can validate our manipulation of bias: do students with no cross-ethnic friends exhibit higher bias in the game? Second, it helps us further unpack bias: what individual characteristics correlate with learning bias?

Formally, for each agent i, we estimate the distribution of bias parameters $h_i(b|y_i, x_i, \theta)$ that is consistent with the estimated parameters θ and the player's individual choices y_i under the choice environment they

[31] A common example Fricker cites is "that the police do not believe [someone] because [they] are black" (p. 1)

faced x_i. The estimated means of this distribution h_i will then give us the individual-level estimates $\hat{b}_{si}, \hat{b}_{di}$. We estimate each h_i using Bayes' rule:

$$h_i(b|y_i, x_i, \theta) = f(b_i|\theta)\frac{\mathbb{P}(y_i|x_i, b_i)}{\mathbb{P}(y_i|x_i, \theta)} \tag{23}$$

Intuitively, we estimate h_i as a posterior distribution. The prior for each h_i is the population-level estimated function $f(b_i|\theta)$. The fraction $\frac{\mathbb{P}(y_i|x_i,b_i)}{\mathbb{P}(y_i|x_i,\theta)}$ represents the (normalized) probability that agent i chose y_i when the parameters are b_i. Multiplying these two distributions gives us the posterior, i.e. $h_i(b_i|y_i, x_i, \theta)$. Practically, for each individual i we take R draws $b_i^r \sim f(b_i|\theta)$. Then, each individual-level bias estimate is a weighted average of these draws:

$$b_{si} = \sum_r w_i^r b_{si}^r \tag{24}$$

$$b_{di} = \sum_r w_i^r b_{di}^r \tag{25}$$

$$\text{where } w_i^r = \frac{\mathbb{P}(y_i|x_i, b_i^r)}{\sum_r \mathbb{P}(y_i|x_i, b_i^r)} \tag{26}$$

In Figure 9, Panel A, we plot the distributions of individual-level biases as estimated from Equations 24 and 25. The distributions seem to match the population-level estimates. In Panels B we split the same-group bias b_s by whether students report any cross-ethnic friends; in Panel C we repeat the same for different-group bias b_d. Recall that in the game we used this indicator as a proxy for different-group bias. Here we find some support for this choice. Visually, the distributions for same-group bias are not clearly distinguishable from each other. A Kolomogorov-Smirnov test cannot reject the null that the distributions are equal ($p = 0.91$). However, the different-group bias distribution seems shifted to the left for the students who report at least one cross-ethnic friend. This implies that students with cross-ethnic friends are incorporate more information from cross-ethnic players. The Kolomogorov-Smirnov test rejects the null that the distributions are the equal ($p = 0.06$), and also rejects the null that the distribution of different-group bias has larger values for those players with at least one cross-ethnic friend ($p = 0.03$).

Finally, in Figure 10 we show correlations between the individual-level estimates of different-group bias and the student's characteristics. Although the regression coefficients are insignificant, the directionality of the

correlations is interesting to note. For example, the more students perceive the different ethnicity in the game to be educated and honest, and the more they speak their language, the lower the bias. On the other end, high perceptions of confidence of the other ethnicity in the game predict larger bias. It is important to note that the largest coefficient in magnitude, and the only significant one is the perceptions of education. That is, the more students perceive the other ethnicity in the game to be educated, the less they discount their information. These correlations do not have a causal interpretation, and can mask complex relationships between identity, perceptions, and bias. More research is needed to better understand the mechanisms through which identity can shape social learning.

6.2 How does integrating classrooms change social learning?

The Mayor of the Municipality in which we conducted our study has talked about a policy of creating one integrated English-language classroom per cohort in the high schools and that they are discussing this policy with the Ministry of Education. In this section, we use model-based counterfactuals to formally explore the implications of the proposed policy. Needless to say that our estimates in this section rely on modeling assumptions and abstractions from complexities in real life. To the extent that such complexities may be important, our estimates should be taken less as precise predictions, and more as suggestive implications. In other settings, researchers can conduct similar exercises under different assumptions that are appropriate in the given context to inform how social learning can change as a consequence of shifts in social networks.

To conduct the counterfactuals we need three ingredients for the model: 1) bias, 2) initial signals, and 3) counterfactual social networks. For the bias we consider the estimated distribution $f(b_i|\theta)$ from the previous section to inform how agents learn from each other. However, students may decrease their bias once they are integrated (Pettigrew, 1998, Mousa, 2020, Lowe, 2021). In a separate study in the same context, we find that cooperative cross-group interactions can decrease bias. We find that when students play video games in mixed pairs, their bias decreases by 0.14 standard deviations compared to same-ethnicity pairs (Anisfeld and Pandiloski, 2024). We therefore consider two bias regimes: one in which bias remains unchanged under integration, i.e. using $f(b_i|\theta)$ from the previous section, and another in which bias decreases by 0.14 standard deviations, i.e. $f(b_i|\theta')$ with $\theta' = (b_s, b_d - 0.14\sigma_d, \sigma_s, \sigma_d)$. A critical assumption here is that the bias

estimated in the lab game is appropriate to apply in other setting and larger networks.[32]

Second, we need prior signals, including an average precision (what fraction of agents get correct signals) as well as the distribution of these signals (how are the correct signals distributed across ethnicities). For the average precision, i.e. the fraction of agents who have a true prior, we use use a grid of 10 equidistant points in the range $\mathbb{P}(\omega = 1) \in [0.5, 1]$. For a given average precision we also consider two possible schedules of distribution. Under uncorrelated priors priors, each ethnic group gets the same precision, such that everyone has the same quality of information regardless of ethnicity. Under correlated priors we distribute the signals unevenly, comparing Albanians on one end with Macedonians and Turks on the other, since the number of Albanians is usually similar to the number of Macedonians and Turks combined. While signals may be unevenly distributed in many ways, focusing on two similarly sized groups allows for simple manipulation and keeps things consistent with the rest of the paper. We set group-level precision as follows. First, we identify whether the number of Albanians or Macedonians and Turks is smaller. The smaller group gets a correct signal with certainty (i.e. everyone has a correct prior), while the larger group gets the rest of the correct signals (if they exist) such that the ethnicity-level average precision adds up to the corresponding precision at the network level. To illustrate, suppose there are 30 Albanians and 35 Macedonians and Turks in one cohort, and suppose that 0.7 of all agents in the network get the correct signal. Under uncorrelated priors, each ethnicity has a precision of 0.7; under correlated priors, Albanians get a precision of 1, while Macedonians and Turks get a precision of 0.4428.

Third, we need a way to construct counterfactual networks. We expand the stochastic block model from Section 3 by allowing that linking probabilities depend not only on ethnicity, but also on classroom assignments. Specifically, we imagine that the network is formed such that each pair meets once and a link forms with one of the following probabilities: $p \in \{p_s^1, p_s^0, p_d^1, p_d^0\}$. The subscript here denotes same- or different-ethnicity, while the superscript is coded as 1 if the pair of students are from the same classroom, and 0 if they are from different classrooms.

As we detail in Section 4, during our baseline survey, we elicited data on social ties from our participants, yielding 12 real-life networks, one for each school-grade level.[33] Using this data on possible and realized

[32] For example, in the game players can observe 2 or 3 people, but in real life students observe more (or fewer) friends. A key assumption in the counterfactuals is that this bias is fixed regardless of the number of observations.

[33] In this section, we define a link between individuals i and j if either i or j listed the other person in at least one of the social

links, we can directly estimate p_s^1, p_s^0, p_d^0. To construct counterfactual networks we need an assumption about what p_d^1 may be after classrooms are integrated, as we do not observe this probability in the data. To retrieve this quantity, we assume that the ratios of own-to-different classroom linking probabilities are the same across ethnicity:

$$\frac{p_s^1}{p_s^0} = \frac{p_d^1}{p_d^0} \implies p_d^1 = p_d^0 \frac{p_s^1}{p_s^0} \tag{27}$$

Finally, we need to specify how the classrooms are assigned in the world of the counterfactual policy. We do this by adding an additional classroom in each cohort, and then assign students in our sample to this classroom with independent probability $1/c$ where c is the total number of classrooms in that cohort. Therefore, the classroom in expectation reflects the ratios of ethnicities in the cohort.[34]

Equipped with an estimation of bias, a range of initial signals, as well as counterfactual networks under the integration policy, we can play out the evolution of learning with the model. Our main outcome in this section is the fraction of agents who guess the correct state of the world at round $t = 10$ of the social learning process. Let learning at the school-cohort level sc as $l_{sc} = \frac{\sum_{i \in N_{sc}} \mathbb{1}[\text{i guessed correctly at t=10}]}{N_{sc}}$. For each school-cohort, we run 1000 iterations within which we pick random bias for each individual in the network, and generate a random counterfactual networks. Then, we average over iterations, yielding $\bar{l}_{sc}$, which we average again across networks $L = \frac{1}{12} \sum_{sc} \bar{l}_{sc}$.

In Figure 11 we present the results of the counterfactual exercises. On the x-axis we plot the precision of the initial signal and on the y-axis we plot the average change in learning L relative to the initial true actions. When priors are uncorrelated, the segregated status-quo network learns better across different initial precision. In this case each ethnic group has the same quality of information, so cross-ethnic links do not increase the flow of good information, while introducing costs in learning due to the higher bias, which results in lower overall learning.[35] On the other hand, when priors are correlated and groups have unequal quality of

survey questions (clique, cross ethnic friends, school help, information). In Table 2, we present summary statistics for the real-life networks.

[34]Note that some learning may be mechanically lower under the counterfactual assumptions. This is because there are fewer students per classroom (since we add an additional classroom), but also because students in the integrated classrooms make fewer friends, since they have fewer co-ethnic friends in their classroom. This leads to a network that has both fewer links and more diverse links. Smaller number of links may therefore mechanically lower learning in the counterfactual networks for reasons beyond diversity of information sources.

[35]This effect is also partly driven by lower number of links in the counterfactual networks. Since every classroom has roughly the same number of students, and linking rates are higher for same-ethnicity peers, the students in the integrated classrooms also have

the signals. Then, the status-quo segregated network can lead to decreases in learning if the initial precision is low. This is because of the echo-chamber effect that makes the group with weak information converge closer to believing the incorrect state – a polarization effect. If the initial precision is low enough, learning can be recovered through integration, since the newly formed links allow true information to pass across the ethnic network clusters. However, if the signal is strong enough, the change in learning is positive, and the status-quo segregated network preserve advantage in learning for the same reasons under uncorrelated priors. This exercise illustrates how integration, bias, and the informational environment matter in shaping social learning. In particular, we illustrate how our modeling approach can be combined with our experimental results to examine how a policy counterfactual of integrated classrooms may shape social learning forces in real life.

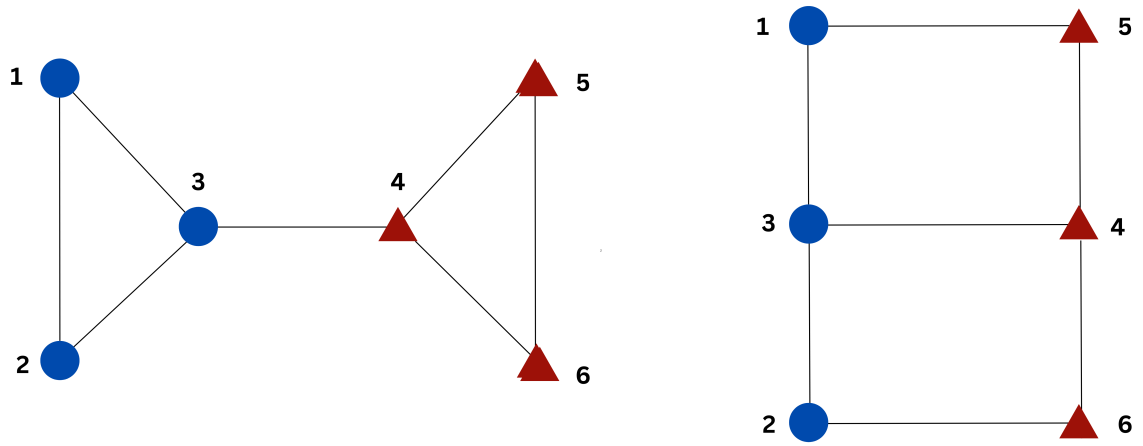
7 Conclusion

In this paper, we explore how integration between different groups affects social learning. We do this in the context of ethnically segregated high schools in Macedonia. We find nuanced effects of integration that depend on both bias and complementarity of initial information. The benefits of integration increase if one group has better information than the other, but decrease if individuals are biased and learn more from their own group. Our finding is analogous to prominent themes in the literature on the pros and cons of diversity, which has focused less attention on social learning.

We also find that students discount information from cross-ethnic peers. The literature that seeks to understand failures of learning and irrational discounting is in its infancy. Our study barely scratches the surface, but it points to a novel direction by suggesting that identity matters. Our finding begets many questions, and more research is needed to understand how identity matters for epistemic failures.

fewer peers under the model.

Tables and Figures



Segregated network **Integrated network**

Figure 1: Observation networks in the game. This figure depicts the networks in the in the lab game. Each network is composed of 6 players from 2 distinct ethnicities. Nodes denote players, and ethnicity is marked by color and shape, so nodes $\{1, 2, 3\}$ represent one ethnicity and nodes $\{4, 5, 6\}$ another one. Each node has a self-link, i.e. observes their own actions. Lines between nodes represent mutual observation, so any two nodes that are linked can see each other's action's during the game.

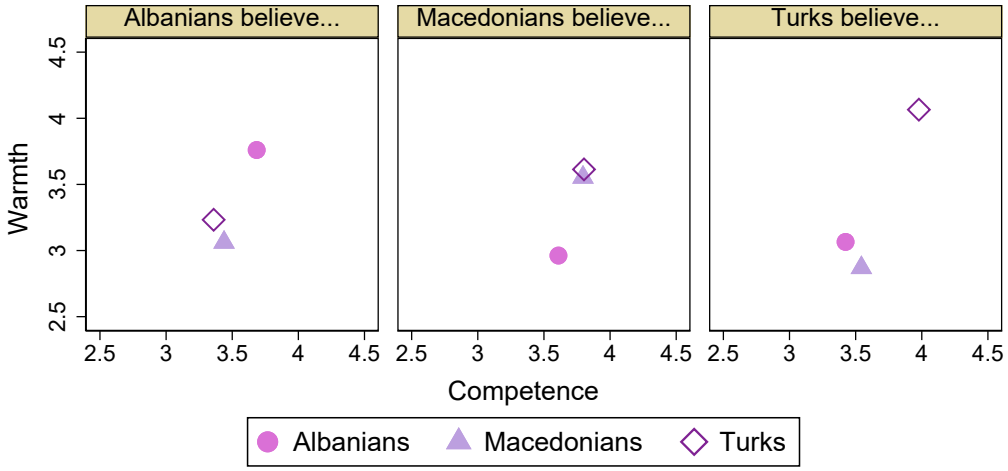
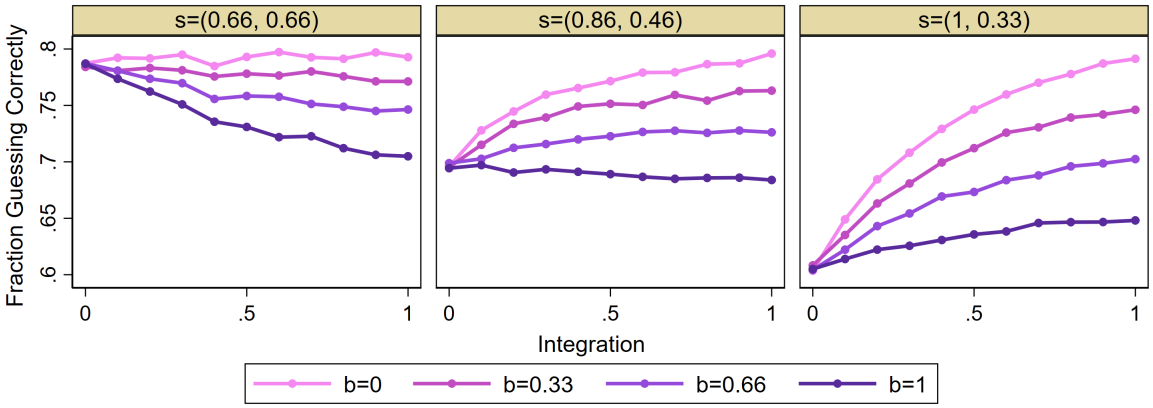
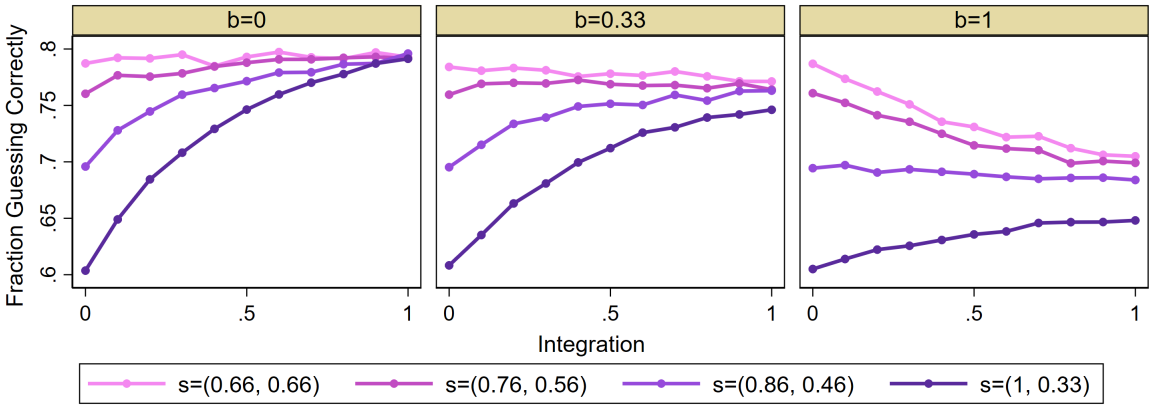


Figure 2: Visualization of Stereotype Content Model questions. This figure visualizes the responses on the Stereotype Content Model questions from Cuddy et al., 2008. On the x-axis we present the Competence stereotypes, which is an average of 4 questions in which participants rank ethnic groups from 1 to 5 on how much they perceive them to be competent, rich, confident, and educated. On the y-axis we present the Warmth stereotypes, which is an average of 2 questions in which participants rank ethnic groups from 1 to 5 on how much they perceive them to be warm and honest. Each panel represents a subset of respondents, with the Albanian respondents shown in the left panel, Macedonian respondents in the middle, and Turkish in the right panel. Their perceptions of different groups are marked by different shapes, triangles representing Macedonians, circles representing Albanians, and diamonds representing Turks.



(a) Panel A: Learning by bias



(b) Panel B: Learning by correlation of prior information

Figure 3: Theoretical results. This figure shows the simulation of the model. On the x-axis we plot *Integration*, defined in the model as $I = \frac{p_d}{p_s}$, and on the y-axis we plot the average learning a, defined as a fraction of agents who correctly guess the state of the world. In Panel A, we present how learning changes over bias. Each line represents a different amount of bias $b = 1 - \frac{1-b_d}{1-b_s}$, and the higher the bias, the darker the shade of the line. Each subpanel presents the results for a different set of correlation of priors. In Panel B, we present how learning changes over correlation of priors. Each line represents a different amount of bias $s = (s_1, s_2)$, and the higher the correlation, the darker the shade of the line. Each subpanel presents the results for a different bias level. In this figure we set the number of individuals in the network to be $n = 100$, the average number of links $d = 5$, same-group bias $b_s = 0$, and the average precision of the signal is 2/3

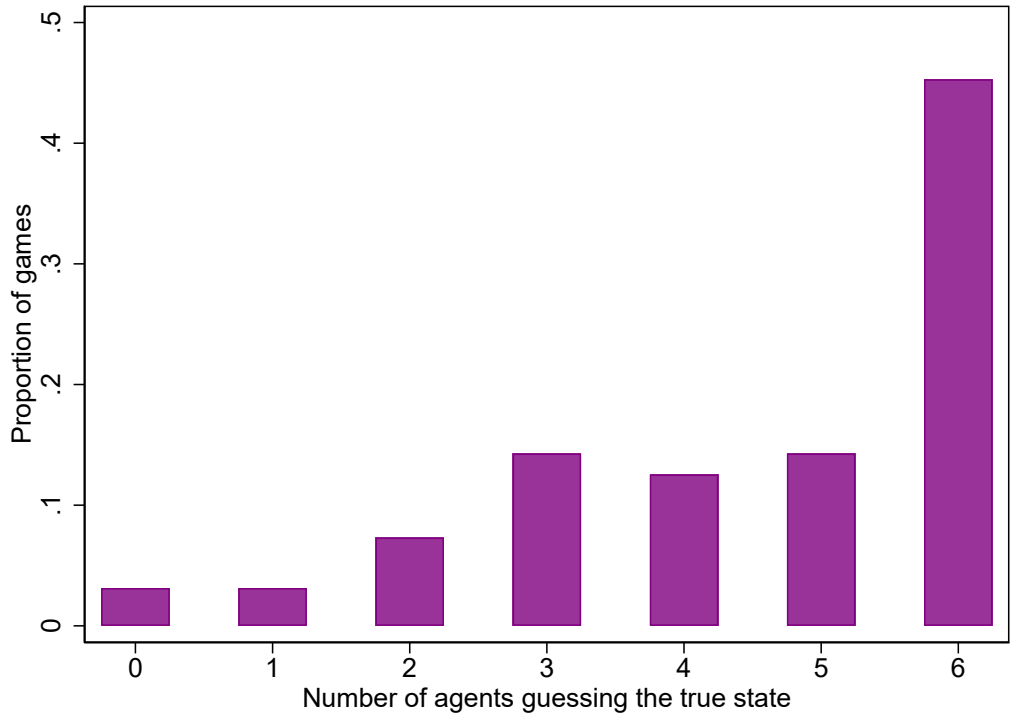
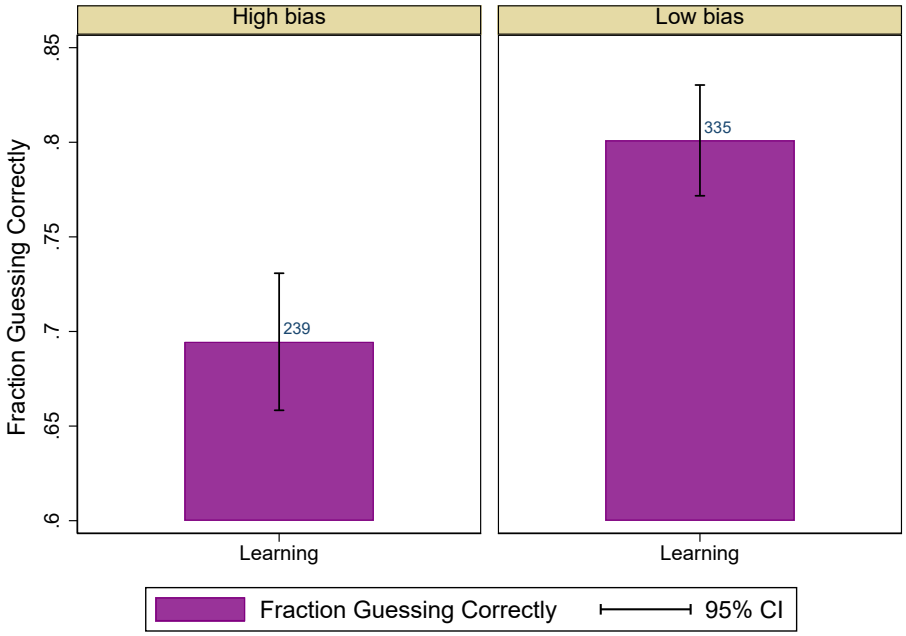
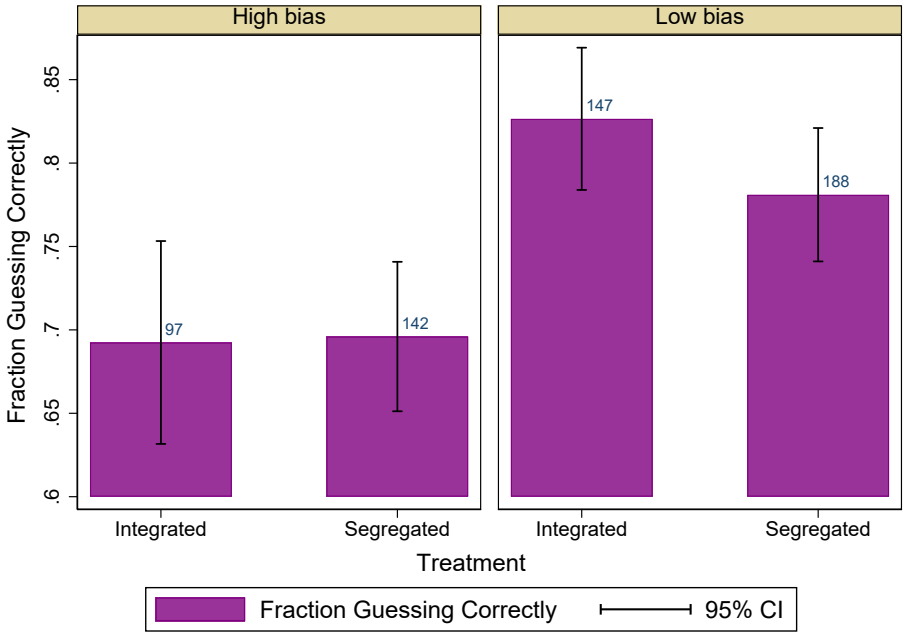


Figure 4: Histogram of learning outcomes. This histogram plots the learning outcome at the network-game level. The *x-axis* presents the number of players that correctly guess the true color in the last round, and the *y-axis* the fraction of network-game outcomes that take that outcome.

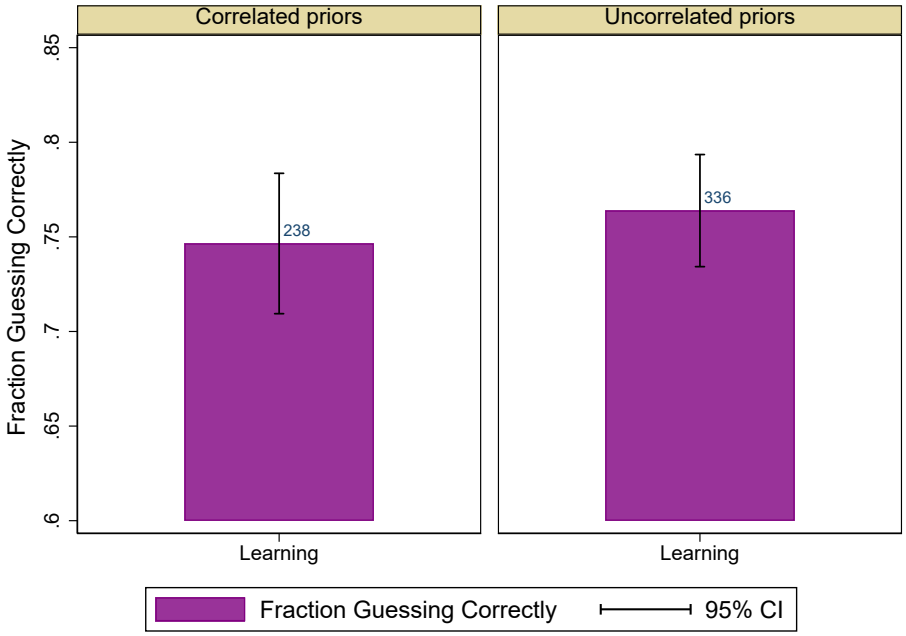


(a) Panel A: Overall learning by bias

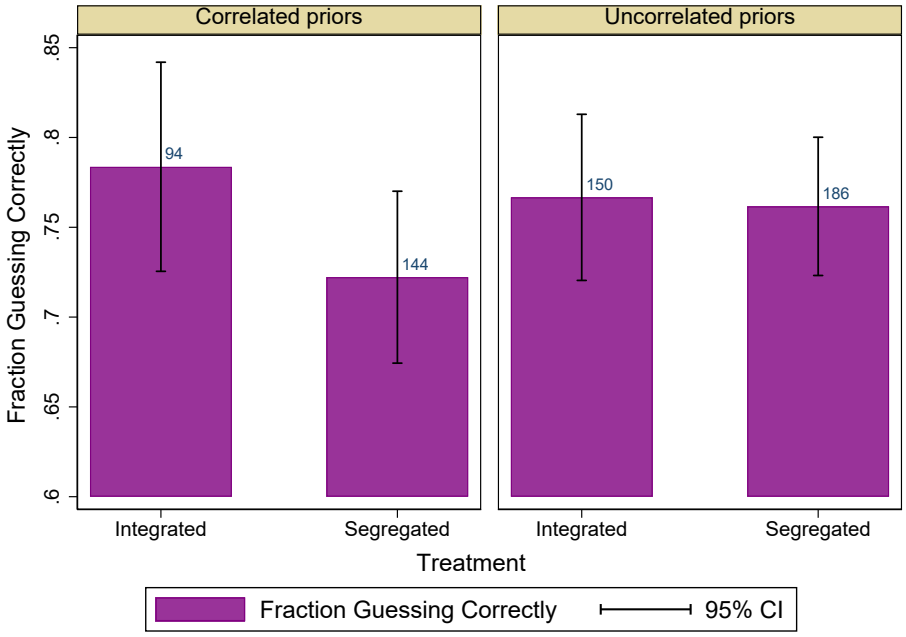


(b) Panel B: Learning by bias, *Segregated* vs. *Integrated* conditions.

Figure 5: Testing Hypothesis 1. This figure plots the mean of learning over network-game observations. Learning is defined as the fraction of the 6 players who correctly guess the color in the last round of the game (N=574). In Panel A, we plot learning by bias. The left panel plots learning for 6-person groups in which half or fewer students have a cross-ethnic friend, while the right panel plots learning for 6-person groups in which more than half of the players have a cross ethnic friendship. For each player, *cross-ethnic friendship* is equal to 1 if the student names at least one friend from a different ethnicity in the baseline social network survey. Panel B is analogous to Panel A, but it additionally splits learning separately for the *Segregated* and *Integrated* conditions. Each figure plots the means, 95% confidence interval, and indicates sample size on top of the bar. We exclude practice rounds.

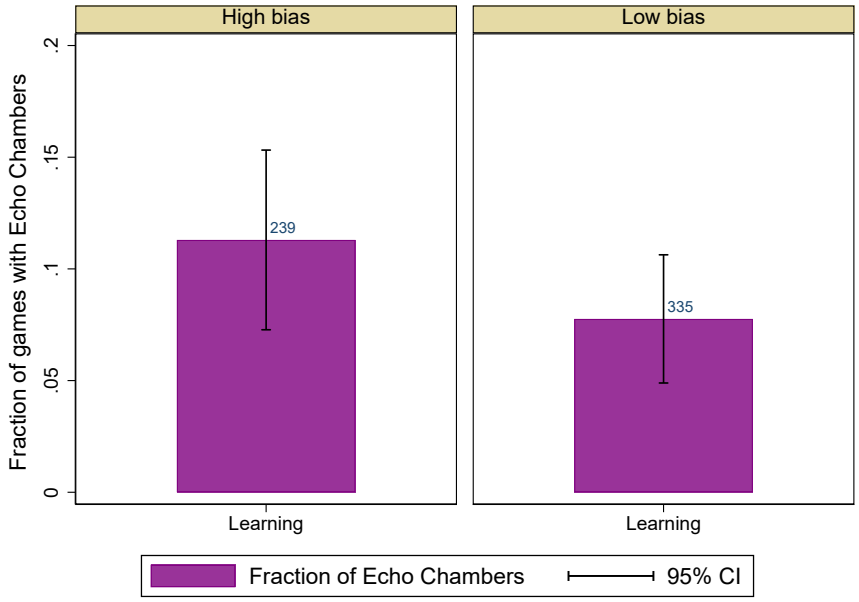


(a) Panel A: Overall learning by correlated priors

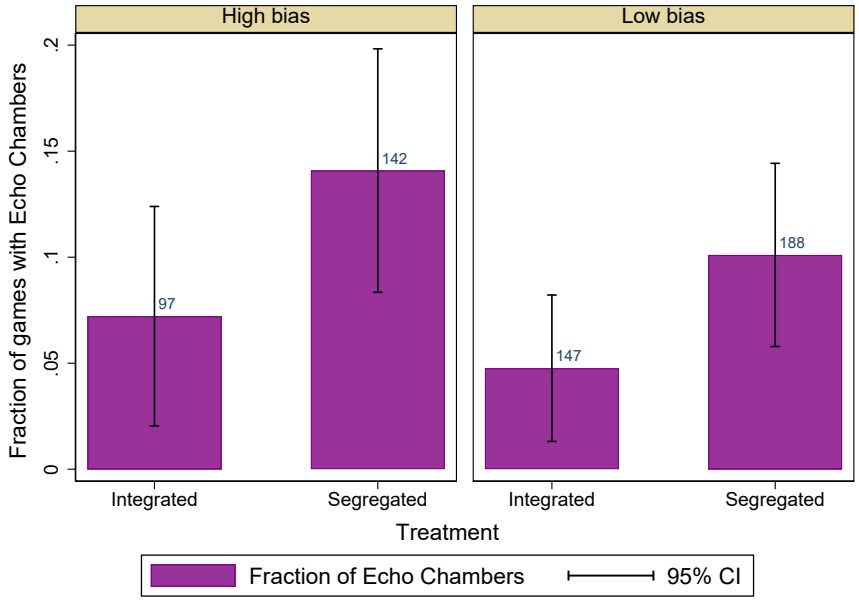


(b) Panel B: Learning by correlated priors, *Segregated* vs. *Integrated* conditions.

Figure 6: Testing Hypothesis 2. This figure plots the mean of learning over network-game observations. Learning is defined as the fraction of the 6 players who correctly guess the color in the last round of the game (N=574). In Panel A, we plot the average true guess by correlated priors. The right panel plots learning for games in which both ethnicities received 2 correct and 1 incorrect signals (uncorrelated priors), while the left panel plots learning for games in which one ethnicity gets 3 correct signals, while the other 2 incorrect and 1 correct signal (correlated priors). In each game signals are distributed to players randomly. Panel B is analogous to Panel A, but additionally splits learning separately for the *Segregated* and *Integrated* conditions. Each figure plots the means, 95% confidence interval, and indicates sample size N on top of the bar. We exclude practice rounds.

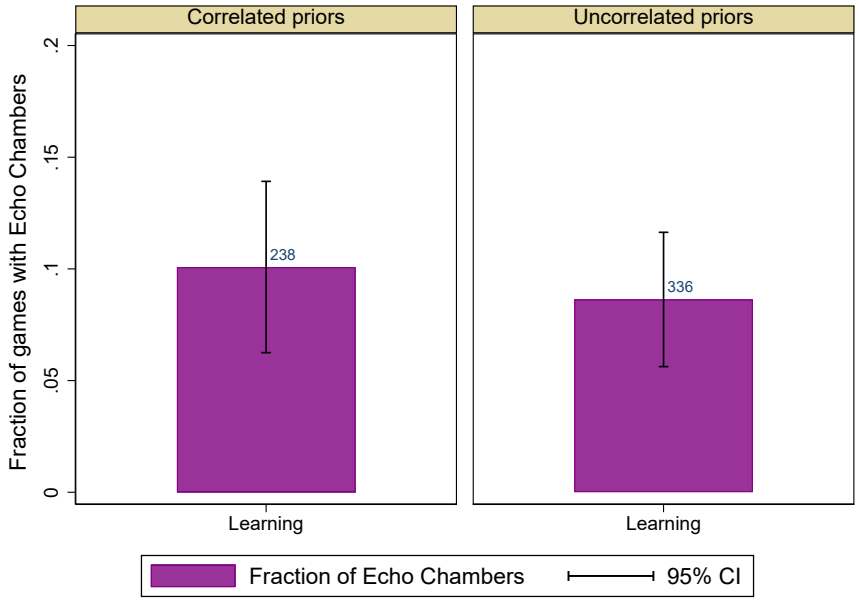


(a) Panel A: Echo chambers by bias

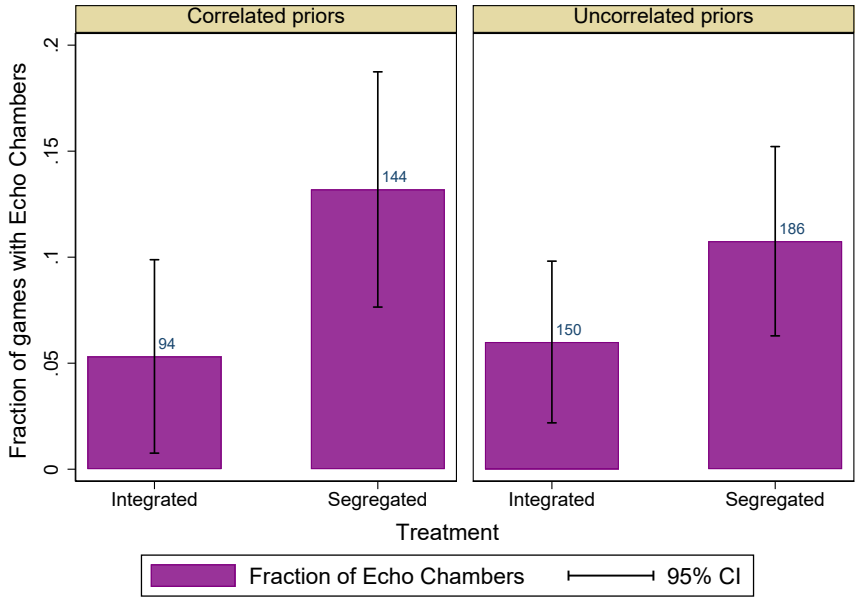


(b) Panel B: Echo chambers by bias, *Segregated* vs. *Integrated* conditions.

Figure 7: Echo chambers and bias. This figure plots the mean of echo chambers over network-game observations. Echo chambers is a dummy variable that takes a value 1 if ethnic groups diverge in opinions: all three students of one ethnicity believe the true state, while all of the three students from the other ethnicity believe the wrong state of the world. In Panel A, we plot echo chambers by bias. The left panel plots learning for 6-person groups in which half or fewer students have a cross-ethnic friend, while the right panel plots echo chambers for 6-person groups in which more than half of the players have a cross ethnic friendship. For each player, *cross-ethnic friendship* is equal to 1 if the student names at least one friend from a different ethnicity in the baseline social network survey. Panel B is analogous to Panel A, but it additionally splits echo chambers separately for the *Segregated* and *Integrated* conditions. Each figure plots the means, 95% confidence interval, and indicates sample size on top of the bar. We exclude practice rounds.

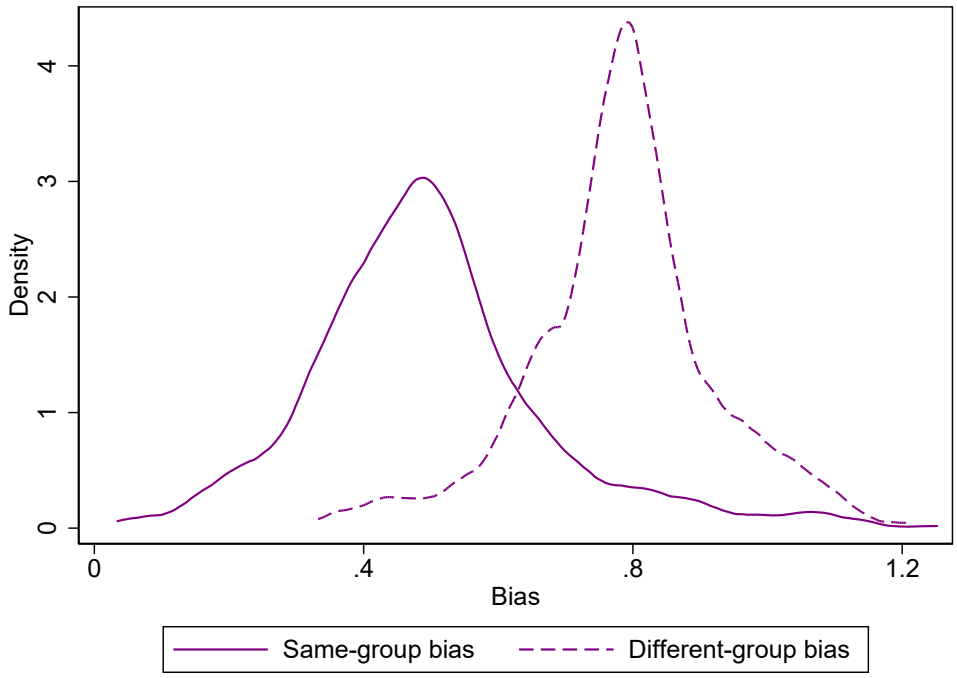


(a) Panel A: Echo chambers by correlated priors

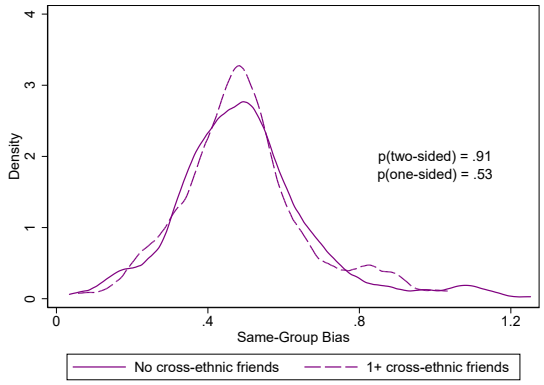


(b) Panel B: Echo chambers by correlated priors, *Segregated* vs. *Integrated* conditions.

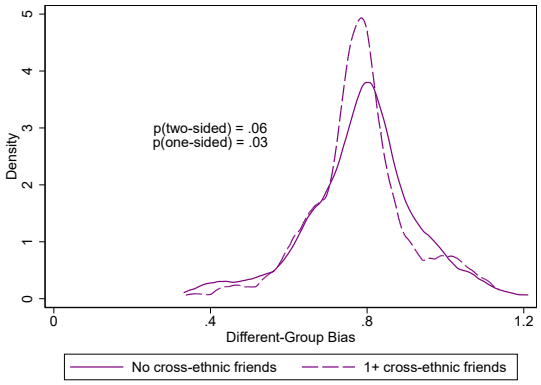
Figure 8: Echo chambers and correlated priors. This figure plots the mean of echo chambers over network-game observations. Echo chambers is a dummy variable that takes a value 1 if ethnic groups diverge in opinions: all three students of one ethnicity believe the true state, while all of the three students from the other ethnicity believe the wrong state of the world. In Panel A, we plot the echo chambers by correlated priors. The right panel plots echo chambers for games in which both ethnicities received 2 correct and 1 incorrect signals (uncorrelated priors), while the left panel plots echo chambers for games in which one ethnicity gets 3 correct signals, while the other 2 incorrect and 1 correct signal (correlated priors). In each game signals are distributed to players randomly. Panel B is analogous to Panel A, but additionally splits echo chambers separately for the *Segregated* and *Integrated* conditions. Each figure plots the means, 95% confidence interval, and indicates sample size N on top of the bar. We exclude practice rounds.



(a) Estimated distributions of bias parameters



(b) Estimated distributions of same-group bias by cross-ethnic friendships



(c) Estimated distributions of different-group bias by cross-ethnic friendships

Figure 9: Distribution of individual-level bias. This figure depicts the distribution of bias estimates at the individual-level, as detailed in Section 6. In Panel A, we show the distribution of same-group bias b_{si} (solid line) and different-group bias b_{di} (dashed line). In Panel B and C we show the distribution of same-group bias b_{si} and different-group bias b_{di} respectively. The distributions in Panels B and C are split separately for students who report no cross-ethnic friends (full line) and those who report at least one cross-ethnic friend (long dashed line). In Panel B and C, we also present the p-value of the Kolomogorov-Smirnov test, from a two-sided test that the distributions are equal, and a one-sided test that the distribution of bias for no cross-ethnic friends is larger.

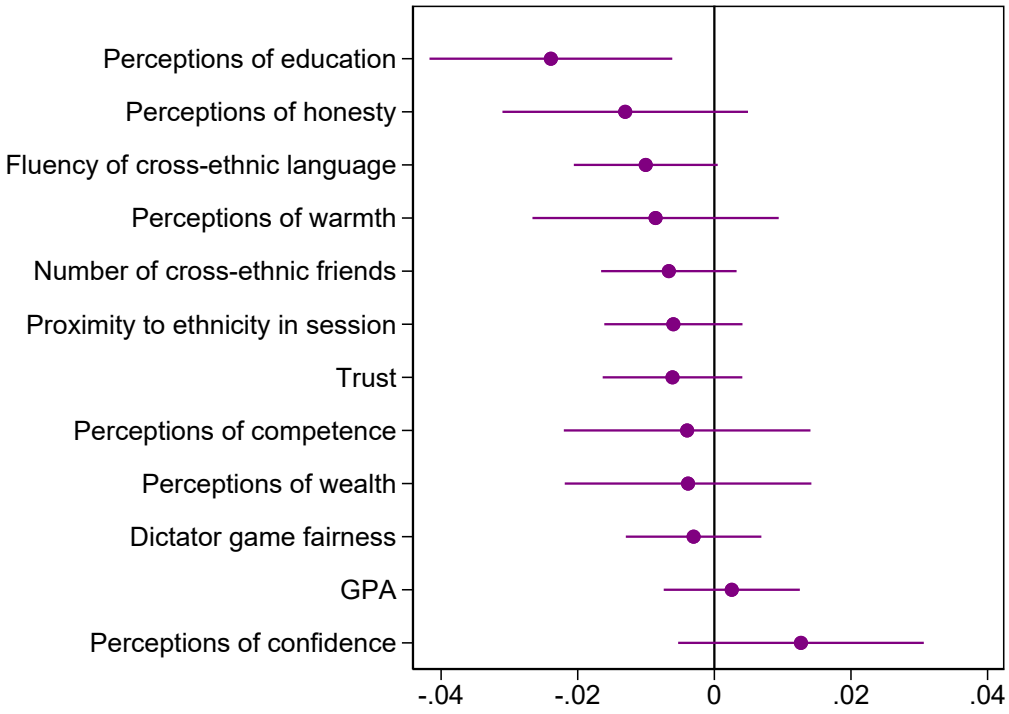


Figure 10: Associations between different-group bias (b_{di}) and individual characteristics. In the first column, we show the coefficients and 90% confidence intervals from a simple regression of different-group bias b_{di} on individual characteristics. The perceptions are responses on how much the players perceive the other ethnicity in the game to be educated, honest, warm, competent, rich, and confident. Fluency of cross-ethnic language is the reported fluency for the language of the other ethnicity in the game. Number of cross ethnic friends is the number of named friends from a different ethnicity at the social network survey. Dictator Game Fairness is coded as 1 if students split cash equally or give more to cross-ethnic students, and 0 otherwise. Trust is a report on generalized trustworthiness in society. GPA is the student's success in school as reported in the administrative school data. Every variable is standardized, so the estimated coefficients are the predicted change in different-group bias when the predictor shifts by one SD.

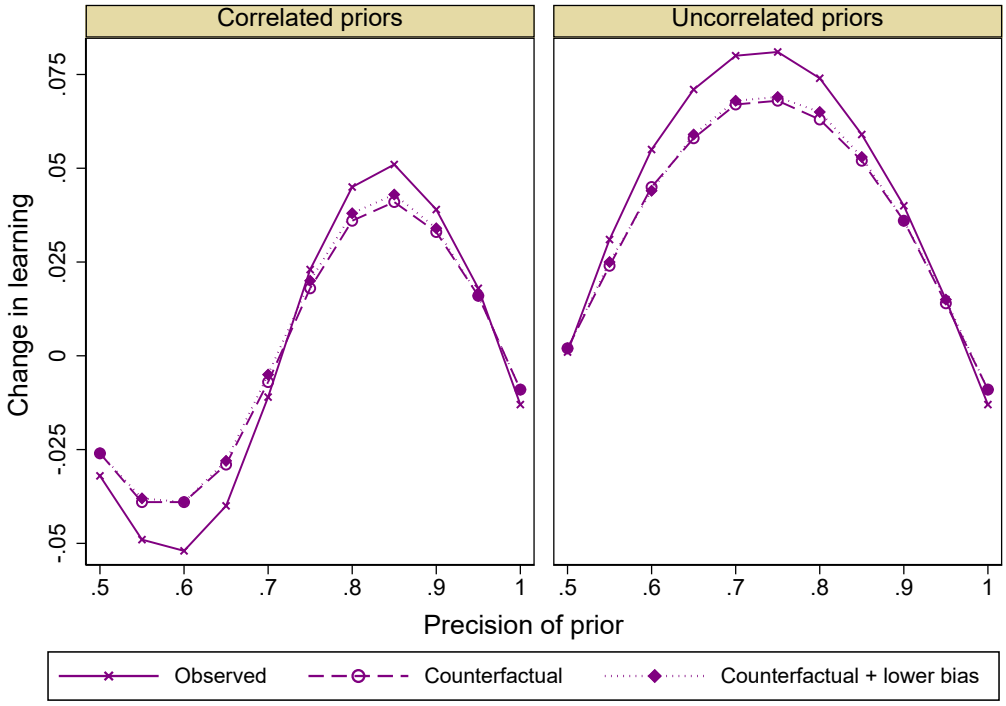


Figure 11: Counterfactuals. This figure plots the results from the counterfactual exercises. On the *x-axis* we plot the precision, i.e. the initial fraction of agents who take the correct action. On the *y-axis* we plot change in learning relative to the initial learning. On the right panel we plot the simulations with uncorrelated priors, such that each ethnicity has the same initial true action, and the left panel presents learning under correlated priors, in which there are differences in the precision of the priors. In each line we plot a different type of network and bias used to simulate social learning. The *Observed* line plots the change in learning using the measured status-quo networks. The *Counterfactual* line plots the change in learning using model-generated networks under the counterfactual scenario, in which one integrated classroom is introduced in each cohort. The *Counterfactual+lower bias*, takes the same counterfactual network, but assumes that integration also lowers bias in learning in the same magnitude as in Anisfeld and Pandiloski, 2024. The network formation model and the assumptions made about counterfactual parameters are detailed in Section 6.2.

Characteristic	N	Mean	SD	Min	Max
Female	527	0.49	0.50	0	1
Age at registration	499	16.31	1.12	13	23
Macedonian	527	0.40	0.49	0	1
Albanian	527	0.41	0.49	0	1
Turkish	527	0.19	0.39	0	1
At least one cross-ethnic friend	527	0.45	0.50	0	1
Dictator Game bias	511	32.63	87.00	-300	500
Fluency in other local languages	511	3.19	2.88	0	10
Proximity to other local ethnicities	511	2.08	1.16	1	7
Generalized Trust	509	5.58	2.44	0	10
At least one parent has college degree	521	0.44	0.50	0	1
Both parents have college degrees	527	0.18	0.39	0	1
Wants to attain graduate degree	515	0.50	0.50	0	1
Grade Point Average (2023-2024)	524	4.54	0.72	2.1	5

Table 1: Summary statistics of individual characteristics. This table shows the characteristics of participants. There are 527 unique participants in this study, though some questions have missing responses since they were not mandatory and/or data is missing for some students. Female is a binary indicator equal to 1 if the enumerator-reported gender is female. Age at registration is derived from the self-reported birth month and year at registration. Macedonian, Albanian, and Turkish are binary variables that equal 1 if the participant is enrolled in a Macedonian, Albanian, or Turkish language instruction classrooms. At least one cross ethnic friend is coded as 1 if the student reports at least one friend from another instruction language in the baseline survey. Dictator game bias is defined as the average amount sent to a co-ethnic minus the average amount sent to a cross-ethnic peer on a Dictator Game. Fluency in other local languages is the average reported fluency (1-10 rank) in local languages that is not the participant's mother tongue. Proximity to other local ethnicities is the average reported proximity on an Inclusion of Others in Self psychometric test. Trust is a self report (0-10 rank) on the general trustworthiness of people in society. Parent education and own educational attainment is measured by a survey report, in which students report the highest educational level each of their parents attained, and the highest educational level they want to attain. GPA comes from administrative school data, and students are graded from 1 (fail) to 5 (excellent).

		Subgraph Network (surveyed students)				
School	Grade	N	Degree	Clustering	Components	Cross links
Gymnasium	1	94	5.13	0.45	6	7.88
	2	84	4.05	0.39	11	9.41
	3	54	2.56	0.41	12	1.45
	4	22	2.55	0.45	4	0.00
Technical	1	70	2.63	0.31	19	4.35
	2	77	3.92	0.40	9	3.97
	3	40	3.30	0.46	6	1.52
	4	52	2.69	0.49	14	0.00
Medical	1	92	4.07	0.35	3	6.95
	2	101	5.41	0.44	2	5.13
	3	85	3.81	0.41	7	4.94
	4	49	2.49	0.48	11	1.64

		Star Network (all students)				
School	Grade	N	Degree	Clustering	Components	Cross links
Gymnasium	1	172	7.31	0.17	3	21.10
	2	205	8.16	0.18	1	12.57
	3	138	10.00	0.16	1	8.90
	4	114	6.54	0.20	3	19.41
Technical	1	252	11.54	0.21	1	9.90
	2	229	8.90	0.21	7	12.00
	3	170	8.50	0.22	6	12.03
	4	71	9.45	0.27	3	17.31
Medical	1	234	9.57	0.18	1	7.23
	2	243	9.43	0.23	1	9.04
	3	199	7.53	0.23	4	9.14
	4	148	7.61	0.21	6	9.28

Table 2: Summary statistics of social networks. This table shows network statistics for the 12 real-life networks. Here we use data from all students with baseline surveys, not just those who have played the game. The Subgraph Network panel shows the data for the social networks that only include surveyed respondents. The Star Network panel includes data on all reported links, even a student who was named as a friend did not participate in the baseline survey. Average degree is the average number of reported links. Clustering represents the probability that two people are linked given that they have a mutual friend. Components is the number of disconnected clusters in the network. Cross links represent the percentage of links in the network that are across individuals of different ethnicities.

	Network-level Learning			
	Hypothesis 1.1		Hypothesis 1.2	
	(1)	(2)	(3)	(4)
High Bias	-0.094*	-0.094*		
	(0.04)	(0.04)		
High Bias X Segregated			0.008	0.014
			(0.04)	(0.04)
Low Bias X Integrated			0.129**	0.134**
			(0.05)	(0.05)
Low Bias X Segregated			0.076	0.077
			(0.05)	(0.05)
Constant	0.643***	0.643***	0.544***	0.542***
	(0.05)	(0.05)	(0.06)	(0.06)
Strata FE	Y	Y	Y	Y
Ethnic pairing FE	N	Y	N	Y
R^2	0.116	0.140	0.121	0.146
N	574	574	574	574
θ	-0.094	-0.094	-0.060	-0.071
$\Pr(\theta > 0)$	0.005***	0.008***	0.079*	0.048**

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

Table 3: Testing Hypothesis 1. This table presents the regressions that test Hypothesis 1. In every column, the outcome is the network-level learning in the last round of each game, measured as the fraction of the 6 students that correctly guessed the color. Columns (1) and (2) test whether bias decreases learning, and map onto the raw means presented in Figure 5 Panel A. Columns (3) and (4) test whether bias increases the costs of integration, and map onto the raw means presented in Figure 5 Panel B. Columns (1) and (3) include school-by-cohort strata fixed effects, while columns (2) and (4) add a fixed effect for the ethnic pairing in the session (Macedonian-Albanian, Albanian-Turkish, Turkish-Macedonian). At the bottom of each column, we show the estimated θ, which theory predicts $\theta \leq 0$. Below it, we present the p-value associated with the one-sided test of the null hypothesis that $\theta > 0$.

	No cross-ethnic friends			Any cross ethnic friend			
	N	Mean	SD	N	Mean	SD	Difference
Dictator Game bias	291	34.71	93.60	220	29.89	77.54	-4.822
Fluency in other local languages	291	2.21	2.48	220	4.48	2.87	2.266***
Proximity to other local ethnicities	291	1.87	1.15	220	2.34	1.12	0.466***
Trust	291	6.71	2.41	220	6.44	2.52	-0.274
At least one parent has college degree	287	0.47	0.50	234	0.41	0.49	-0.060
Both parents have college degrees	291	0.20	0.40	236	0.17	0.37	-0.034
Wants to attain graduate degree	286	0.52	0.50	229	0.47	0.50	-0.046
GPA	288	4.62	0.66	236	4.44	0.78	-0.183***

Table 4: What's in a "type"? This table shows the characteristics of participants with no cross-ethnic friends and at least one cross-ethnic friend, and the difference on these characteristics between these two groups (coefficient from a regression of given characteristic on any cross friend indicator). Dictator game bias is defined as the average amount sent to a co-ethnic minus the average amount sent to a cross-ethnic peer on a Dictator Game. Fluency in other local languages is the average reported fluency (1-10 rank) in local languages, excluding the participant's mother tongue. Proximity to other local ethnicities is the average reported proximity on an Inclusion of Others in Self psychometric test. Trust is a self report (0-10 rank) on the general trustworthiness of people in society. Parent education and own educational attainment is measured by a survey report, in which students report the highest educational level each of their parents attained, and the highest educational level they want to attain. GPA data comes from administrative data sets in the schools.

	Network-level Learning			
	Hypothesis 2.1		Hypothesis 2.2	
	(1)	(2)	(3)	(4)
Correlated Signals	-0.019	-0.024		
	(0.02)	(0.02)		
Uncorrelated Priors X Integrated			0.004	0.006
			(0.02)	(0.02)
Correlated Priors X Segregated			-0.043	-0.047
			(0.03)	(0.03)
Correlated Priors X Integrated			0.021	0.017
			(0.04)	(0.04)
Constant	0.617***	0.619***	0.617***	0.618***
	(0.02)	(0.02)	(0.03)	(0.03)
Strata FE	Y	Y	Y	Y
Ethnic pairing FE	N	Y	N	Y
R^2	0.094	0.124	0.099	0.129
N	574	574	574	574
θ	-0.019	-0.024	-0.060	-0.058
$\Pr(\theta > 0)$	0.216	0.163	0.053*	0.058*

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

Table 5: Testing Hypothesis 2. This table presents the regressions that test Hypothesis 2. In every column, the outcome is the network-level learning in the last round of each game, measured as the fraction of the 6 students that correctly guessed the color. Columns (1) and (2) test whether correlated priors decrease learning, and map onto the raw means presented in Figure 6 Panel A. Columns (3) and (4) test whether correlated priors increases the benefit of integration, and map onto the raw means presented in Figure 6 Panel B. Columns (1) and (3) include school-by-cohort strata fixed effects, while columns (2) and (4) add a fixed effect for the ethnic pairing in the session (Macedonian-Albanian, Albanian-Turkish, Turkish-Macedonian). At the bottom of each column, we present the estimated θ which theory predicts $\theta \leq 0$. Below it, we present the p-value associated with the one-sided test of the null hypothesis that $\theta > 0$.

	Echo Chambers			
	(1)	(2)	(3)	(4)
High Bias	0.007	0.033		
	(0.03)	(0.03)		
High Bias X Segregated			0.071*	0.073*
			(0.03)	(0.03)
Low Bias X Integrated			0.001	-0.020
			(0.04)	(0.03)
Low Bias X Segregated			0.060*	0.035
			(0.03)	(0.03)
Constant	-0.002	-0.011	-0.036	-0.020
	(0.01)	(0.01)	(0.02)	(0.02)
Strata FE	Y	Y	Y	Y
Ethnic pairing FE	N	Y	N	Y
R^2	0.068	0.075	0.080	0.086
N	574	574	574	574
θ	0.007	0.033	0.013	0.018
$\Pr(\theta < 0)$	0.402	0.140	0.380	0.335

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

Table 6: Echo chambers and bias. This table presents analogous regressions to those that test Hypothesis 1 in Table 3. In every column, the outcome is whether the learning resulted in echo chambers in the last round of each game. Echo chambers is a dummy variable that takes a value 1 if ethnic groups diverge in opinions: all three students of one ethnicity believe the true state, while all of the three students from the other ethnicity believe the wrong state of the world. Columns (1) and (2) test whether bias increases echo chambers, and map onto the raw means presented in Figure 7 Panel A. Columns (3) and (4) test whether segregation leads to more echo chambers when bias is high, and map onto the raw means presented in Figure 7 Panel B. Columns (1) and (3) include school-by-cohort strata fixed effects, while columns (2) and (4) add a fixed effect for the ethnic pairing in the session (Macedonian-Albanian, Albanian-Turkish, Turkish-Macedonian). At the bottom of each column, we show the estimated θ, which our hypotheses imply $\theta \geq 0$. Below it, we present the p-value associated with the one-sided test of the null hypothesis that $\theta < 0$.

	Echo Chambers			
	(1)	(2)	(3)	(4)
Correlated Signals	0.015	0.014		
	(0.03)	(0.03)		
Uncorrelated Priors X Integrated			-0.045	-0.045
			(0.03)	(0.03)
Correlated Priors X Segregated			0.031	0.030
			(0.04)	(0.04)
Correlated Priors X Integrated			-0.059*	-0.059*
			(0.03)	(0.03)
Constant	-0.005	-0.005	0.014	0.015
	(0.01)	(0.01)	(0.02)	(0.02)
Strata FE	Y	Y	Y	Y
Ethnic pairing FE	N	Y	N	Y
R^2	0.069	0.073	0.082	0.086
N	574	574	574	574
θ	0.015	0.014	0.045	0.044
$\Pr(\theta < 0)$	0.288	0.300	0.175	0.179

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

Table 7: Echo chambers and correlated priors. This table presents analogous regressions to those that test Hypothesis 2 in Table 5. In every column, the outcome is whether the learning resulted in echo chambers in the last round of each game. Echo chambers is a dummy variable that takes a value 1 if ethnic groups diverge in opinions: all three students of one ethnicity believe the true state, while all of the three students from the other ethnicity believe the wrong state of the world. Columns (1) and (2) test whether correlated priors increase echo chambers, and map onto the raw means presented in Figure 8 Panel A. Columns (3) and (4) test whether segregation leads to more echo chambers under correlated priors, and map onto the raw means presented in Figure 8 Panel B. Columns (1) and (3) include school-by-cohort strata fixed effects, while columns (2) and (4) add a fixed effect for the ethnic pairing in the session (Macedonian-Albanian, Albanian-Turkish, Turkish-Macedonian). At the bottom of each column, we present the estimated θ which our hypotheses imply $\theta \geq 0$. Below it, we present the p-value associated with the one-sided test of the null hypothesis that $\theta < 0$.

Parameter	Estimate (std. err.)
Same-group bias mean (b_s)	0.50
	(0.08)
Different-group bias mean (b_d)	0.79
	(0.07)
Same-group bias SD (σ_s)	0.35
	(0.08)
Different-group bias SD (σ_d)	0.34
	(0.05)

Table 8: Estimated biases. This table presents the maximum simulated likelihood estimates of population-level distribution of biases. The first two rows present the estimated means for the same-group and different-group bias distribution, respectively, and the last two rows show the estimated standard deviations. The mixed logit estimation procedure is detailed in Section 6. Standard errors are derived via bootstrap sampling, in which we re-estimated the parameters 500 times (estimates stabilize at around 300 samples).

References

Acemoglu, D., Dahleh, M. A., Lobel, I., & Ozdaglar, A. (2011). Bayesian Learning in Social Networks. *The Review of Economic Studies*, *78*(4), 1201–1236. https://doi.org/10.1093/restud/rdr004

Alesina, A., Baqir, R., & Easterly, W. (1999). Public Goods and Ethnic Divisions*. *The Quarterly Journal of Economics*, *114*(4), 1243–1284. https://doi.org/10.1162/003355399556269

Alesina, A., & Ferrara, E. L. (2005). Ethnic Diversity and Economic Performance. *Journal of Economic Literature*, *43*(3), 762–800. https://doi.org/10.1257/002205105774431243

Alesina, A., Harnoss, J., & Rapoport, H. (2016). Birthplace diversity and economic prosperity. *Journal of Economic Growth*, *21*(2), 101–138. https://doi.org/10.1007/s10887-016-9127-6

Anisfeld, A., & Pandiloski, P. (2024). Language, Identity, and Ethnic Cohesion in Schools. *Working Paper*.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*(4), 596–612. https://doi.org/10.1037/0022-3514.63.4.596

Bahar, D., & Rapoport, H. (2018). Migration, Knowledge Diffusion and the Comparative Advantage of Nations. *The Economic Journal*, *128*(612), F273–F305. https://doi.org/10.1111/ecoj.12450

Banerjee, A., Breza, E., Chandrasekhar, A. G., & Mobius, M. (2021). Naïve Learning with Uninformed Agents. *American Economic Review*, *111*(11), 3540–3574. https://doi.org/10.1257/aer.20181151

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013). The Diffusion of Microfinance. *Science*, *341*(6144), 1236498. https://doi.org/10.1126/science.1236498

Banerjee, A., Iyer, L., & Somanathan, R. (2005). History, Social Divisions, and Public Goods in Rural India. *Journal of the European Economic Association*, *3*(2/3), 639–647.

Bayer, P., Ross, S. L., & Topa, G. (2008). Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes. *Journal of Political Economy*, *116*(6), 1150–1196. https://doi.org/10.1086/595975

Beaman, L., & Magruder, J. (2012). Who Gets the Job Referral? Evidence from a Social Networks Experiment. *The American Economic Review*, *102*(7), 3574–3593.

Beaman, L. A. (2012). Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the U.S. *The Review of Economic Studies*, *79*(1), 128–161. https://doi.org/10.1093/restud/rdr017

Berman, N., Couttenier, M., Rohner, D., & Thoenig, M. (2017). This Mine is Mine! How Minerals Fuel Conflicts in Africa. *The American Economic Review*, *107*(6), 1564–1610.

Besley, T., & Persson, T. (2011). The Logic of Political Violence*. *The Quarterly Journal of Economics*, *126*(3), 1411–1445. https://doi.org/10.1093/qje/qjr025

Bramoullé, Y., Djebbari, H., & Fortin, B. (2020). Peer Effects in Networks: A Survey. *Annual Review of Economics*, *12*(Volume 12, 2020), 603–629. https://doi.org/10.1146/annurev-economics-020320-033926

Bursztyn, L., Ederer, F., Ferman, B., & Yuchtman, N. (2014). Understanding Mechanisms Underlying Peer Effects: Evidence From a Field Experiment on Financial Decisions. *Econometrica*, *82*(4), 1273–1301. https://doi.org/10.3982/ECTA11991

Cai, J., De Janvry, A., & Sadoulet, E. (2015). Social Networks and the Decision to Insure. *American Economic Journal: Applied Economics*, *7*(2), 81–108.

Chandrasekhar, A. G. (2016, April). Econometrics of Network Formation. In *The Oxford Handbook of the Economics of Networks* (pp. 302–357). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199948277.013.21

Chandrasekhar, A. G., Larreguy, H., & Xandri, J. P. (2020). Testing Models of Social Learning on Networks: Evidence From Two Experiments. *Econometrica*, *88*(1), 1–32. https://doi.org/10.3982/ECTA14407

Conley, T. G., & Udry, C. R. (2010). Learning about a New Technology: Pineapple in Ghana. *American Economic Review*, *100*(1), 35–69. https://doi.org/10.1257/aer.100.1.35

Conlon, J. J., Mani, M., Rao, G., Ridley, M., & Schilbach, F. (n.d.). Learning in the Household.

Conlon, J. J., Mani, M., Rao, G., Ridley, M. W., & Schilbach, F. (2022, August). Not Learning from Others. https://doi.org/10.3386/w30378

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008, January). Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map. In *Advances in Experimental Social Psychology* (pp. 61–149, Vol. 40). Academic Press. https://doi.org/10.1016/S0065-2601(07)00002-0

Damm, A. P. (2009). Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-Experimental Evidence. *Journal of Labor Economics*, *27*(2), 281–314. https://doi.org/10.1086/599336

DeGroot, M. H. (1974). Reaching a Consensus. *Journal of the American Statistical Association*, *69*(345), 118–121. https://doi.org/10.2307/2285509

Duflo, E., & Saez, E. (2002). Participation and investment decisions in a retirement plan: The influence of colleagues' choices. *Journal of Public Economics*.

Duflo, E., & Saez, E. (2003). The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment. *The Quarterly Journal of Economics*, *118*(3), 815–842.

Dupas, P. (2014). Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence from a Field Experiment. *Econometrica*, *82*(1), 197–228.

Esteban, J., & Ray, D. (2011). Linking Conflict to Inequality and Polarization. *American Economic Review*, *101*(4), 1345–1374. https://doi.org/10.1257/aer.101.4.1345

Foster, A. D., & Rosenzweig, M. R. (1995). Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture. *Journal of Political Economy*, *103*(6), 1176–1209.

French Jr., J. R. P. (1956). A formal theory of social power. *Psychological Review*, *63*(3), 181–194. https://doi.org/10.1037/h0046123

Fricker, M. (2007, June). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198237907.001.0001

Glaeser, E. L., Sacerdote, B., & Scheinkman, J. A. (1996). Crime and Social Interactions*. *The Quarterly Journal of Economics*, *111*(2), 507–548. https://doi.org/10.2307/2946686

Goel, D., & Lang, K. (2019). Social Ties and the Job Search of Recent Immigrants. *ILR Review*, *72*(2), 355–381. https://doi.org/10.1177/0019793917729350

Golub, B., & Sadler, E. (n.d.). LEARNING IN SOCIAL NETWORKS.

Golub, B., & Jackson, M. O. (2010). Naïve Learning in Social Networks and the Wisdom of Crowds. *American Economic Journal: Microeconomics*, *2*(1), 112–149. https://doi.org/10.1257/mic.2.1.112

Golub, B., & Jackson, M. O. (2012). How Homophily Affects the Speed of Learning and Best-Response Dynamics. *The Quarterly Journal of Economics*, *127*(3), 1287–1338. https://doi.org/10.1093/qje/qjs021

Golub, B., & Sadler, E. (2016, June). Learning in Social Networks. In *The Oxford Handbook of the Economics of Networks* (Yann Bramoullé, Andrea Galeotti, and Brian W. Rogers).

Granovetter, M. (1983). The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory*, *1*, 201–233. https://doi.org/10.2307/202051

Harary, F. (1959). A criterion for unanimity in French's theory of social power. In *Studies in social power* (pp. 168–182). Univer. Michigan.

Hjort, J. (2014). Ethnic Divisions and Production in Firms *. *The Quarterly Journal of Economics*, *129*(4), 1899–1946. https://doi.org/10.1093/qje/qju028

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton University Press.

Jha, S. (2013). Trade, Institutions, and Ethnic Tolerance: Evidence from South Asia. *American Political Science Review*, *107*(4), 806–832. https://doi.org/10.1017/S0003055413000464

Khwaja, A. I. (2009). Can good projects succeed in bad communities? *Journal of Public Economics*, *93*(7), 899–916. https://doi.org/10.1016/j.jpubeco.2009.02.010

Kremer, M., & Miguel, E. (2007). The Illusion of Sustainability*. *The Quarterly Journal of Economics*, *122*(3), 1007–1065. https://doi.org/10.1162/qjec.122.3.1007

Lazarsfeld, P. F., Berelson, B., & Gaudet, H. (1944). *The people's choice*. Duell, Sloan & Pearce.

Lazear, E. P. (1999). Globalisation and the Market for Team-Mates. *The Economic Journal*, *109*(454), C15–C40.

Lowe, M. (2021). Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration. *American Economic Review*, *111*(6), 1807–1844. https://doi.org/10.1257/aer.20191780

Massey, C., & Wu, G. (2005). Detecting Regime Shifts: The Causes of Under- and Overreaction. *Management Science*, *51*(6), 932–947. https://doi.org/10.1287/mnsc.1050.0386

Meyer, C. D. (2023, January). *Matrix Analysis and Applied Linear Algebra, Second Edition*. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611977448

Michalopoulos, S., & Papaioannou, E. (2016). The Long-Run Effects of the Scramble for Africa. *The American Economic Review*, *106*(7), 1802–1848.

Miguel, E., & Gugerty, M. K. (2005). Ethnic diversity, social sanctions, and public goods in Kenya. *Journal of Public Economics*, *89*(11), 2325–2368. https://doi.org/10.1016/j.jpubeco.2004.09.004

Miller, G., & Mobarak, A. M. (n.d.). GENDER DIFFERENCES IN PREFERENCES, INTRA-HOUSEHOLD EXTERNALITIES, AND THE LOW DEMAND FOR IMPROVED COOKSTOVES.

Mobius, M., Phan, T., & Szeidl, A. (2015, March). Treasure Hunt: Social Learning in the Field. https://doi.org/10.3386/w21014

Mobius, M., & Rosenblat, T. (2014). Social Learning in Economics. *Annual Review of Economics*, *6*(Volume 6, 2014), 827–847. https://doi.org/10.1146/annurev-economics-120213-012609

Montalvo, J. G., & Reynal-Querol, M. (2005). Ethnic Polarization, Potential Conflict, and Civil Wars. *The American Economic Review*, *95*(3), 796–816.

Montalvo, J. G., & Reynal-Querol, M. (2021). Ethnic Diversity and Growth: Revisiting the Evidence. *The Review of Economics and Statistics*, *103*(3), 521–532. https://doi.org/10.1162/rest_a_00901

Mousa, S. (2020). Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq. *Science*, *369*(6505), 866–870. https://doi.org/10.1126/science.abb3153

Munshi, K. (2004). Social learning in a heterogeneous population: Technology diffusion in the Indian Green Revolution. *Journal of Development Economics*, *73*(1), 185–213. https://doi.org/10.1016/j.jdeveco.2003.03.003

Nathan, M. (2016). Ethnic diversity and business performance: Which firms? Which cities? *Environment and Planning A: Economy and Space*, *48*(12), 2462–2483. https://doi.org/10.1177/0308518X16660085

Oster, E., & Thornton, R. (2012). Determinants of Technology Adoption: Peer Effects in Menstrual Cup Take-Up. *Journal of the European Economic Association*, *10*(6), 1263–1293. https://doi.org/10.1111/j.1542-4774.2012.01090.x

Ottaviano, G. I., & Peri, G. (2006). The economic value of cultural diversity: Evidence from US cities. *Journal of Economic Geography*, *6*(1), 9–44. https://doi.org/10.1093/jeg/lbi002

Peri, G. (2012). The Effect Of Immigration On Productivity: Evidence From U.S. States. *The Review of Economics and Statistics*, *94*(1), 348–358. https://doi.org/10.1162/REST_a_00137

Pettigrew, T. F. (1998). INTERGROUP CONTACT THEORY. *Annual Review of Psychology*, *49*(Volume 49, 1998), 65–85. https://doi.org/10.1146/annurev.psych.49.1.65

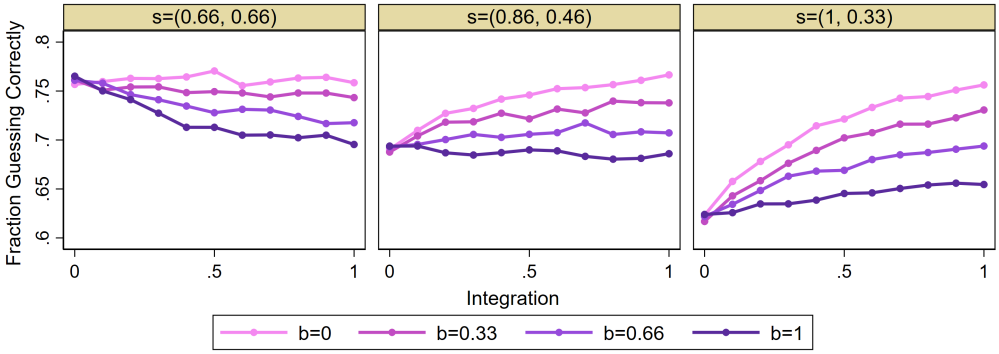
Train, K. E. (2009). *Discrete Choice Methods with Simulation* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511805271

Weizsäcker, G. (2010). Do We Follow Others When We Should? A Simple Test of Rational Expectations. *American Economic Review*, *100*(5), 2340–2360. https://doi.org/10.1257/aer.100.5.2340

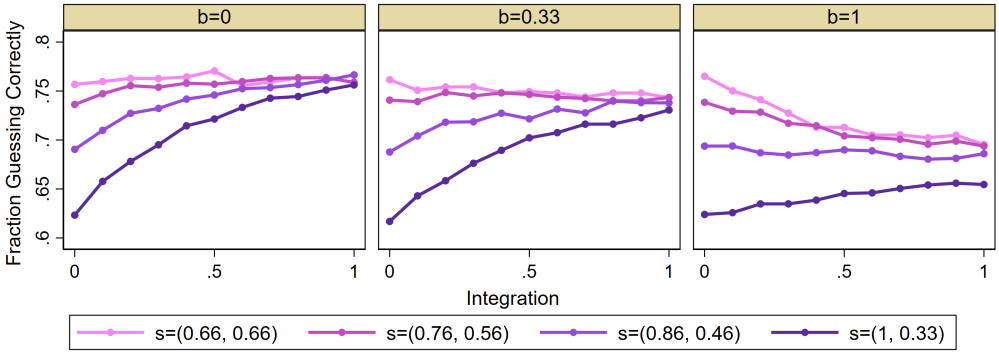
Zelizer, A. (2019). Is Position-Taking Contagious? Evidence of Cue-Taking from Two Field Experiments in a State Legislature. *American Political Science Review*, *113*(2), 340–352. https://doi.org/10.1017/S0003055419000078

A Theoretical Appendix

A.1 Additional simulations

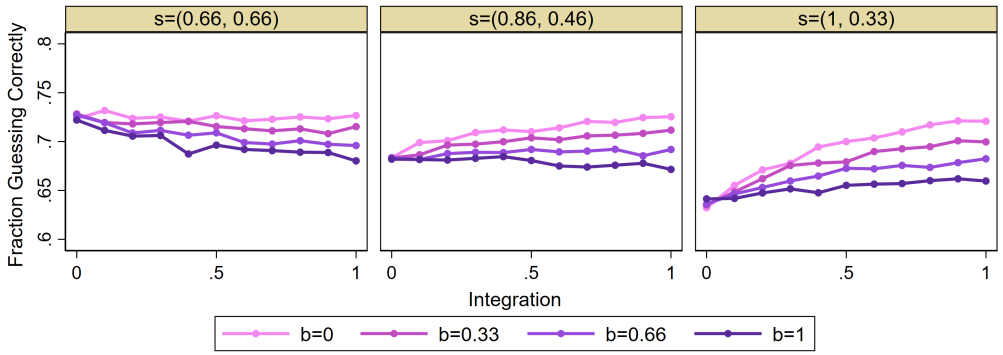


(a) Panel A: Learning by bias

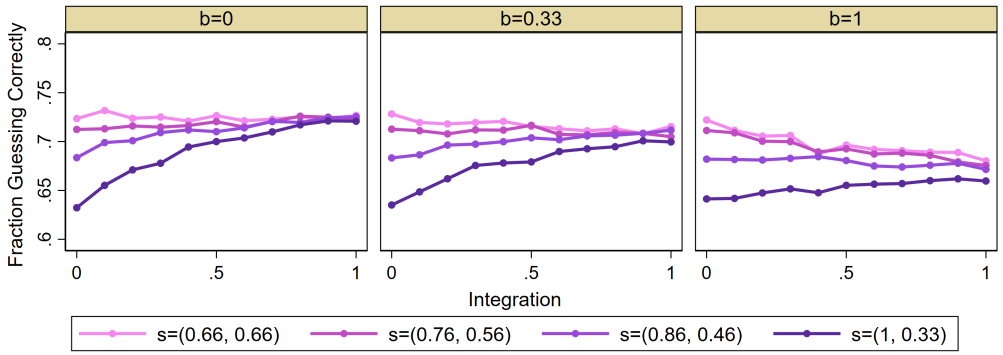


(b) Panel B: Learning by correlation of prior information

Figure A1: Theoretical results, robustness with $b_s = 0.25$. This figure shows the simulation results of the model. It is analogous to Figure 3, but instead of setting $b_s = 0$, we set $b_s = 0.25$. On the x-axis we plot *Integration*, defined in the model as $I = \frac{p_s}{p_d}$, and on the y-axis we plot the average learning a, defined as a fraction of agents who correctly guess the state of the world. In Panel A, we present how learning changes over bias. Each line represents a different amount of bias $b = 1 - \frac{1-b_d}{1-b_s}$, and the higher the bias, the darker the shade of the line. Each subpanel presents the results for a different set of correlation of priors. In Panel B, we present how learning changes over correlation of priors. Each line represents a different amount of bias $s = (s_1, s_2)$, and the higher the correlation, the darker the shade of the line. Each subpanel presents the results for a different bias level. In this figure we set the number of individuals in the network to be $n = 100$, the average number of links $d = 5$, same-group bias $b_s = 0$, and the average precision of the signal is 2/3

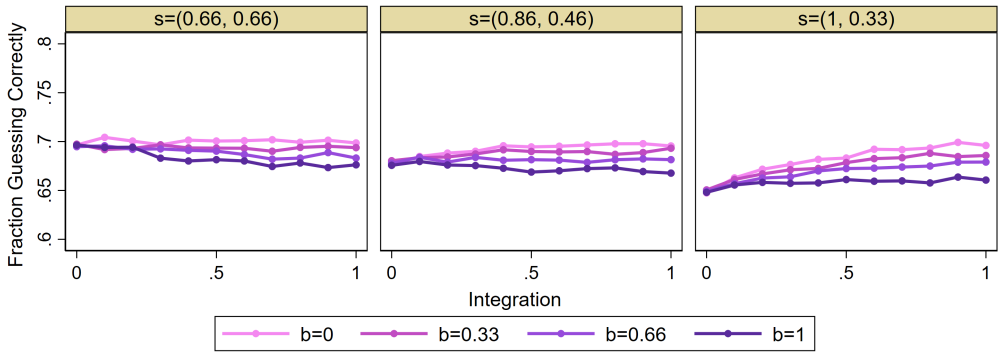


(a) Panel A: Learning by bias

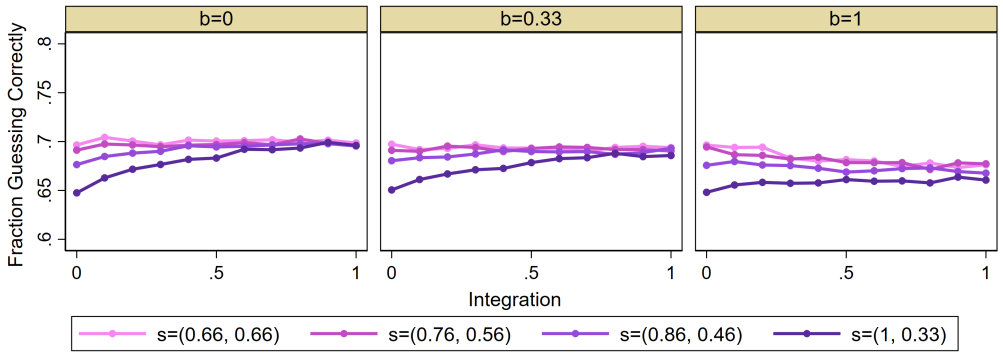


(b) Panel B: Learning by correlation of prior information

Figure A2: Theoretical results, robustness with $b_s = 0.5$. This figure shows the simulation results of the model. It is analogous to Figure 3, but instead of setting $b_s = 0$, we set $b_s = 0.5$. On the x-axis we plot *Integration*, defined in the model as $I = \frac{p_s}{p_d}$, and on the y-axis we plot the average learning a, defined as a fraction of agents who correctly guess the state of the world. In Panel A, we present how learning changes over bias. Each line represents a different amount of bias $b = 1 - \frac{1-b_d}{1-b_s}$, and the higher the bias, the darker the shade of the line. Each subpanel presents the results for a different set of correlation of priors. In Panel B, we present how learning changes over correlation of priors. Each line represents a different amount of bias $s = (s_1, s_2)$, and the higher the correlation, the darker the shade of the line. Each subpanel presents the results for a different bias level. In this figure we set the number of individuals in the network to be $n = 100$, the average number of links $d = 5$, same-group bias $b_s = 0$, and the average precision of the signal is 2/3

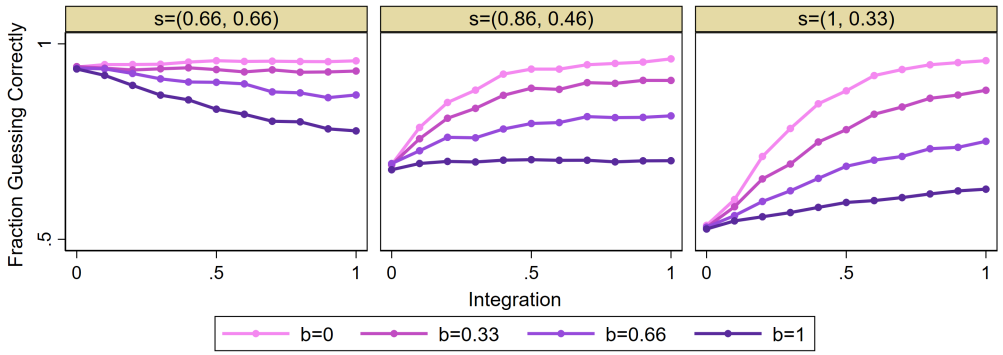


(a) Panel A: Learning by bias

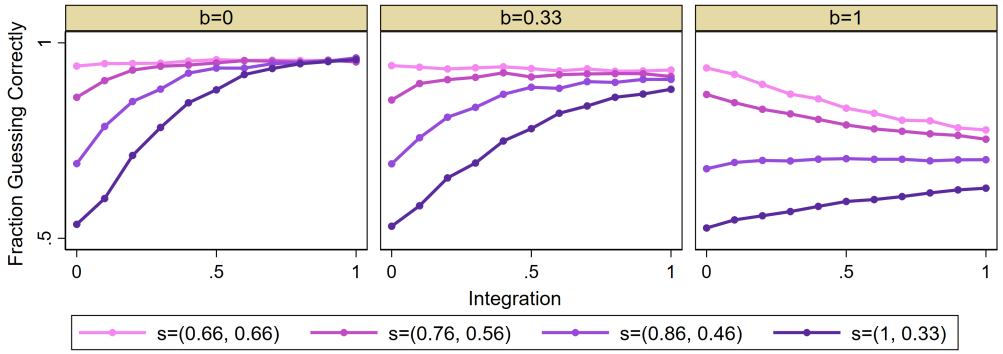


(b) Panel B: Learning by correlation of prior information

Figure A3: Theoretical results, robustness with $d = 2$. This figure shows the simulation results of the model. It is analogous to Figure 3, but instead of setting $d = 5$, we set $d = 2$. On the x-axis we plot *Integration*, defined in the model as $I = \frac{p_s}{p_d}$, and on the y-axis we plot the average learning a, defined as a fraction of agents who correctly guess the state of the world. In Panel A, we present how learning changes over bias. Each line represents a different amount of bias $b = 1 - \frac{1-b_d}{1-b_s}$, and the higher the bias, the darker the shade of the line. Each subpanel presents the results for a different set of correlation of priors. In Panel B, we present how learning changes over correlation of priors. Each line represents a different amount of bias $s = (s_1, s_2)$, and the higher the correlation, the darker the shade of the line. Each subpanel presents the results for a different bias level. In this figure we set the number of individuals in the network to be $n = 100$, the average number of links $d = 5$, same-group bias $b_s = 0$, and the average precision of the signal is 2/3

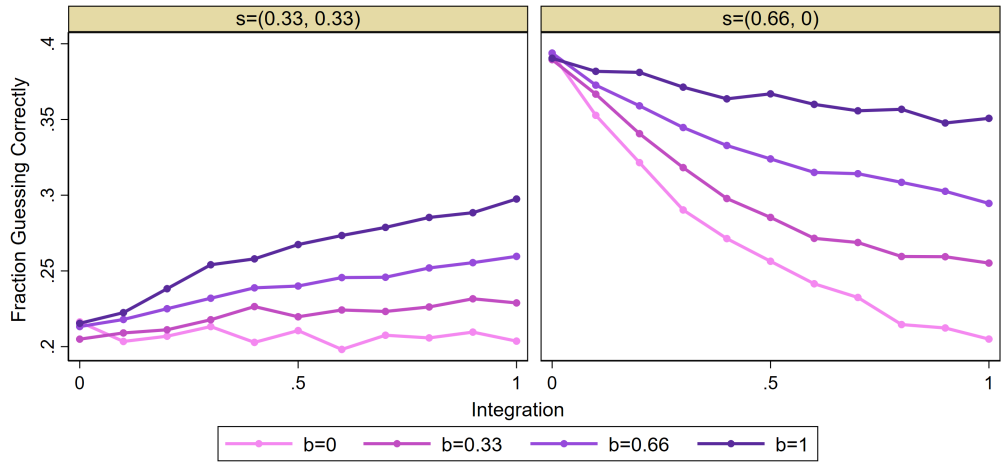


(a) Panel A: Learning by bias

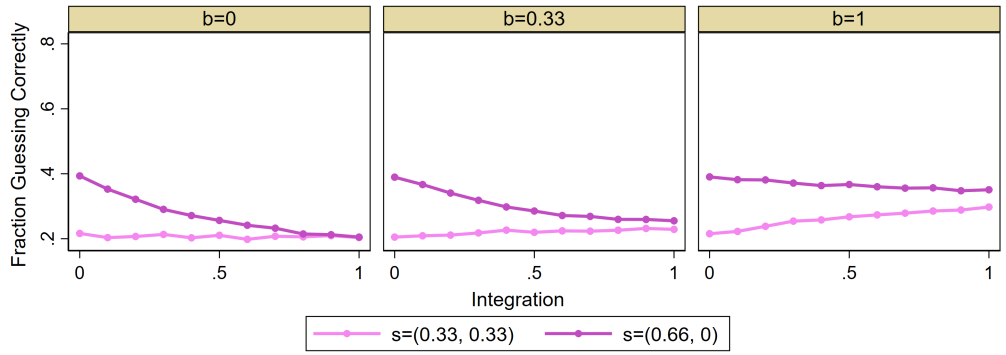


(b) Panel B: Learning by correlation of prior information

Figure A4: Theoretical results, robustness with $d = 10$. This figure shows the simulation results of the model. It is analogous to Figure 3, but instead of setting $d = 5$, we set $d = 2$. On the x-axis we plot *Integration*, defined in the model as $I = \frac{p_s}{p_d}$, and on the y-axis we plot the average learning a, defined as a fraction of agents who correctly guess the state of the world. In Panel A, we present how learning changes over bias. Each line represents a different amount of bias $b = 1 - \frac{1-b_d}{1-b_s}$, and the higher the bias, the darker the shade of the line. Each subpanel presents the results for a different set of correlation of priors. In Panel B, we present how learning changes over correlation of priors. Each line represents a different amount of bias $s = (s_1, s_2)$, and the higher the correlation, the darker the shade of the line. Each subpanel presents the results for a different bias level. In this figure we set the number of individuals in the network to be $n = 100$, the average number of links $d = 5$, same-group bias $b_s = 0$, and the average precision of the signal is 2/3



(a) Panel A: Learning by bias



(b) Panel B: Learning by correlation of prior information

Figure A5: Theoretical results, substitute priors. This figure shows the simulation results of the model. It is analogous to Figure 3, but instead of setting an average precision of 0.66 it sets the average precision to 0.33. This scenario is outside the empirical experiment, which always administers signals with high precision, yet it is theoretically interesting. This setting essentially flips the theoretical hypotheses. On the x-axis we plot *Integration*, defined in the model as $I = \frac{p_s}{p_d}$, and on the y-axis we plot the average learning a, defined as a fraction of agents who correctly guess the state of the world. In Panel A, we present how learning changes over bias. Each line represents a different amount of bias $b = 1 - \frac{1-b_d}{1-b_s}$, and the higher the bias, the darker the shade of the line. Each subpanel presents the results for a different set of correlation of priors. In Panel B, we present how learning changes over correlation of priors. Each line represents a different amount of bias $s = (s_1, s_2)$, and the higher the correlation, the darker the shade of the line. Each subpanel presents the results for a different bias level. In this figure we set the number of individuals in the network to be $n = 100$, the average number of links $d = 5$, same-group bias $b_s = 0$, and the average precision of the signal is 2/3

A.2 The case of asympotic learning with continuous updating

Our modelling approach here mirrors Golub and Jackson, 2012, a seminal theoretical study of how segregation slows learning in social networks (henceforth GJ). They too employ a stochastic block model as a network formation process as well as a DeGroot model as a social learning process. We departs in two key ways. First, we extend their model by introducing a bias parameter b, such that individuals are allowed to be biased against individuals from different groups; in GJ, $b = 0$ and so individuals do not have any bias against other types. Second, GJ consider a continuous action space in an asymptotic learning framework, that is, as society grows to infinity $(n \to \infty)$ and with an updating rule analogous to Equation (4) but without the signum function (given the continuous action space). Building on their work, we generate two propositions that mirror the two main hypotheses in the previous subsection. Namely, both bias and correlated priors slow asymptotic learning. In the end of this subsection, we discuss why certain features of the modelling approach in GJ preclude us from examining the precise frictions we are interested in.

Let's consider this extended model with bias. Continuous updating and asymptotic learning come with some convenient results from the theory of Markov chains. Importantly, with continuous updating, convergence to the same belief is guaranteed if the social network is strongly connected and aperiodic.[36] The assumptions that GJ impose on the networks[37] ensure there is always convergence to the same action in the network.[38] Since disagreement is precluded, the outcome of interest is the *speed* of learning, captured by the following measure of consensus time *CT*. For a given $\epsilon > 0$ and an interaction matrix $\mathbf{G}$, define:

$$CT(\epsilon, \mathbf{G}) = \sup_{\mathbf{a}} \min\{t : CD(t; \mathbf{G}, \mathbf{a}) < \epsilon\} \tag{28}$$

where $CD(t; \mathbf{G}, \mathbf{a}) = ||\mathbf{T}(\mathbf{G})^t\mathbf{a} - \mathbf{T}(\mathbf{G})^\infty\mathbf{a}||^2_{\mathbf{s}(\mathbf{G})}$ is the consensus distance, which is a weighted sum of distances between the belief at time t and the asymptotic belief, with weights $\mathbf{s}(\mathbf{G}) = (\frac{d_1}{D(\mathbf{G})}, ..., \frac{d_n}{D(\mathbf{G})})$, where $d_i(\mathbf{G}) = \sum_{j \in N} G_{ij}$ the weighted degree of agent i and $D(\mathbf{G}) = \sum_i d_i(\mathbf{G})$ as the sum of weighted

[36] Strongly connected means that the there is a path between any i and j in the network, and aperiodic means that the lowest common divisor of all cycles in the network is 1, which is true if for at least one i, $A_{ii} = 1$). See Golub and Sadler, n.d. for a more detailed discussion, and Meyer, 2023 for technical details.

[37] The focus on networks with four features: sufficient density, no vanishing groups, interior homophily, and comparable densities between groups

[38] Additional assumptions are needed to ensure that the convergence is to the true state of the world: 1) the majority of agents get the correct signal, 2) no subnetworks have outsized influence, 3) no subnetworks ignores most of the network. (Golub and Jackson, 2010)

degrees in the network.

Define an 2×2 matrix of *effective connectivity* $\mathbf{F}(b, p_s, p_d)$ with characteristic entries

$$F_{zq}(b, p_s, p_d) = \begin{cases} \frac{p_s}{p_s + p_d(1-b)}, & \text{if } z = q \\ \frac{p_d(1-b)}{p_s + p_d(1-b)}, & \text{if } z \neq q \end{cases} \tag{29}$$

The matrix $\mathbf{F}$ has the same dimension as the number of groups, and each entry F_{zq} measures the fraction of weighted degrees that agents of group z have with agents of group q. The main contribution of GJ is to show that the second eigenvalue[39] of the matrix $\mathbf{F}$ bounds the convergence time in the DeGroot model, i.e., for any $\epsilon > 0$, there is a large enough n such that:

$$\frac{(1-\epsilon)}{2}\left(\frac{log(n)}{log\left(\frac{1}{|\lambda_2(\mathbf{F})|}\right)}\right) \leq CT(\epsilon; \mathbf{G}) \leq 2(1+\epsilon)\left(\frac{log(n)}{log\left(\frac{1}{|\lambda_2(\mathbf{F})|}\right)}\right) \tag{30}$$

Essentially, the speed of learning is sandwiched by functions of the second eigenvalue of the matrix $\mathbf{F}$. The larger the second eigenvalue, the slower the learning. The proofs for the main results in GJ are technical, and we refer the interested reader to their paper for more details. The main results translate in our case, since the arguments rely on large-n dynamics as the realized links in the network approach the expected links, which is true in the model extension with bias presented here. The following proposition relies on the connection between second eigenvalue of $\mathbf{F}$ and speed of learning.

Proposition 1. Bias slows learning:

$$\frac{\partial \lambda_2(\mathbf{F}(b, p_s, p_d))}{\partial b} > 0 \tag{31}$$

Proof. First we show that the second eigenvalue in this case is identical to:

$$\lambda_2(\mathbf{F}(b, p_s, p_d)) = \frac{p_s - p_d(1-b)}{p_s + p_d(1-b)} \tag{32}$$

[39] If we calculate the eigenvalues associated with the matrix $\mathbf{F}$ and ordered them by absolute value from largest to smallest, the second eigenvalue is the second on that list. The largest eigenvalue of a stochastic matrix is unique and is unity, so the second eigenvalue is always smaller than 1in absolute value.

Note that we can decompose

$$\mathbf{F} = \frac{p_d(1-b)}{p_s + p_d(1-b)}\mathbf{O} + \frac{p_s - p_d(1-b)}{p_s + p_d(1-b)}\mathbf{I} \tag{33}$$

where $\mathbf{O}$ is a 2×2 matrix of ones and $\mathbf{I}$ is the 2×2 identity matrix. Since $\mathbf{O}$ and $\mathbf{I}$ commute (are simultaneously diagonalizable), the eigenvalues of F are the sum of the respective eigenvalues of $\mathbf{O}$ and $\mathbf{I}$. Since $\mathbf{O}$ has rank 1, it has only one non-zero eigenvalue, so the second eigenvalue of this matrix is zero. The second matrix is diagonal, so its eigenvalues are equal to its diagonal entries, so all the eigenvalues are equal to $\frac{p_s - p_d(1-b)}{p_s + p_d(1-b)}$. Adding the two vectors of eigenvalues, we get that the second largest eigenvalue of $\mathbf{F}$ is $\frac{p_s - p_d w}{p_s + (m-1)p_d w}$.

Taking derivative of this quantity we obtain:

$$\frac{1}{\partial b}\left(\frac{p_s - p_d(1-b)}{p_s + p_d(1-b)}\right) = \frac{p_s p_d}{(p_s + p_d(1-b))^2} > 0 \tag{34}$$

□

This characterization of the eigenvalue some intuitive interpretations. First, the speed of convergence is a function of both segregation and bias. Just like in GJ, the second eigenvalue λ_2 increases in segregation, through either increasing p_s or decreasing p_d, slow convergence; in our extended model, bias also slows convergence.

Importantly, in this extended model, the second eigenvalue is not only a measurement of segregation, as is the case in GJ with $b = 0$. Rather here the second eigenvalue captures someas a measurement of "effective connectivity", that is a function of both segregation and bias. To illustrate the consequences of this re-interpretation, consider a scenario in which a social planner increases cross-group linkages. In the initial segregated network the parameters describing society are p_s, p_d^s, b^s, while after integration the society is characterized by p_s, p_d^i, b^i, such that $p_d^i > p_d^s$. In the model of GJ, $b^i = b^s = 0$, and the integration would always lead to a faster convergence. However, if this integration also sufficiently *increases* bias, such that $b^i > 1 - (1-b^s)\frac{p_d^s}{p_d^i}$, then integration will *slow* convergence.

We can also analyze how the initial distribution of signals impacts the time to consensus. GJ develop a "representative agent theorem", which shows that asymptotically we can treat each group in the network

as one individual. Intuitively, as $n \to \infty$, the amount of information each agent receives grows large, and convergence within type happens in only 1 period. From the first period on, the speed of learning in the asymptotic case depends on how different groups are connected and what their priors are. Recall that s_1, s_2 represent the average precision for each group, and let $\overline{a}_k(t)$ represent the average action of group k at time t. Then, we can employ the representative agent theorem to show that the disagreement between the two groups at time t is:

Proposition 2. (Golub and Jackson, 2012) The disagreement between the groups at time t is proportional to the second eigenvalue and the difference in priors. Specifically:

$$\overline{a}_1(t) - \overline{a}_2(t) = \lambda_2^t(s_1 - s_2) \tag{35}$$

Proof. This proof uses the "representative agent theorem", that implies that we can asymptotically treat each type of agent as a single individual. Essentially, this means focusing on the expected links to agents of each type. We will first show that:

$$\overline{a}_i(t) = \frac{1}{2}(1 + \lambda_2^t)s_i + \frac{1}{2}(1 - \lambda_2^t)s_{-i} \tag{36}$$

First, note that $\overline{a}_i(0) = s_i$, and that $\frac{1}{2}(1 + \lambda_2) = \frac{p_s}{p_s + p_d(1-b)}$ and $\frac{1}{2}(1 - \lambda_2) = \frac{p_d(1-b)}{p_s + p_d(1-b)}$, which is straightforward to check. The proof proceeds by induction. For $t = 1$:

$$\begin{aligned}
\overline{a}_i(1) &= \frac{p_s}{p_s + p_d(1-b)}\overline{a}_i(0) + \frac{p_d(1-b)}{p_s + p_d(1-b)}\overline{a}_{-i}(0) && (37)\\
&= \frac{1}{2}(1 + \lambda_2)s_i + \frac{1}{2}(1 - \lambda_2)s_{-i} && (38)
\end{aligned}$$

Now, assume this is true for $\overline{a}_i(t)$. Then for $t + 1$ we have:

$$\begin{aligned}
\overline{a}_i(t+1) &= \frac{p_s}{p_s + p_d(1-b)}\overline{a}_i(t) + \frac{p_d(1-b)}{p_s + p_d(1-b)}\overline{a}_{-i}(t) && (39)\\
&= \frac{1}{2}(1 + \lambda_2)\left(\frac{1}{2}(1 + \lambda_2^{t-1})s_i + \frac{1}{2}(1 - \lambda_2^{t-1})s_{-i}\right) && (40)\\
&+ \frac{1}{2}(1 - \lambda_2)\left(\frac{1}{2}(1 + \lambda_2^{t-1})s_{-i} + \frac{1}{2}(1 - \lambda_2^{t-1})s_i\right) && (41)\\
&= \frac{1}{2}(1 + \lambda_2^t)s_i + \frac{1}{2}(1 - \lambda_2^t)s_{-i} && (42)
\end{aligned}$$

Which proves the claim. Calculating the difference yields:

$$\begin{aligned} \overline{a}_1(t) - \overline{a}_2(t) &= \left(\frac{1}{2}(1+\lambda_2^t)s_1 + \frac{1}{2}(1-\lambda_2^t)s_2\right) - \left(\frac{1}{2}(1+\lambda_2^t)s_2 + \frac{1}{2}(1-\lambda_2^t)s_1\right) \quad (43) \\ &= \lambda_2^t(s_1 - s_2) \quad (44) \end{aligned}$$

□

There are two main implication of this proposition. First, it shows that an increase in correlation of priors ($|s_1 - s_2|$) slows learning. While the disagreement still tends to zero as t increases[40], it can persist for longer as $s_1 - s_2$ grows. Second, it shows that if the groups have exactly the same priors $s_1 = s_2$, segregation does not matter asymptotically, since convergence happens in one period. Intuitively, if each group has the same prior, as number of agents grows to infinity, agents can learn just as well from their own group as they would from the other group.

Proposition 1 and 2 show how bias and correlated priors each slow learning in the case of asymptotic learning with continuous updating. However, this approach cannot generate disagreements (failure of convergence), or the *interactions* between segregation, bias, and correlated priors that we generate in the previous subsection. That is, under this framework, if priors are not uncorrelated, segregation always harms the speed of learning. In the previous subsection we showed that segregation can *increase* learning, even for small bias. The asymptotic model cannot generate these frictions since the number of links each individual has grows infinite. With this abundance of information, there are no trade-offs in how links are distributed across types. These trade offs are key in generating the patters from our simulations.

[40] Since $|\lambda_2| < 1$ in any stochastic matrix

B Empirical Appendix

B.1 Placebo results with anonymous players

	Network-level Learning			
	Hypothesis 1.1		Hypothesis 1.2	
	(1)	(2)	(3)	(4)
High Bias	-0.085*	-0.062		
	(0.04)	(0.04)		
High Bias X Segregated			-0.003	-0.005
			(0.03)	(0.03)
Low Bias X Integrated			0.072	0.047
			(0.04)	(0.05)
Low Bias X Segregated			0.102*	0.080
			(0.04)	(0.05)
Constant	0.648***	0.640***	0.563***	0.578***
	(0.05)	(0.04)	(0.06)	(0.05)
Strata FE	Y	Y	Y	Y
Ethnic pairing FE	N	Y	N	Y
R-sqr	0.095	0.105	0.096	0.107
N	544	544	544	544
Theta	-0.085	-0.062	0.033	0.039
pval	0.012**	0.067*	0.769	0.804

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

Table A1: Robustness for Hypothesis 1, placebo. This table presents the regressions that test Hypothesis 1, only for the games in which names were hidden (anonymous play). In every column, the outcome is the network-level learning in the last round of each game, measured as the fraction of the 6 students that correctly guessed the color. Columns (1) and (2) test whether bias decreases learning. Columns (3) and (4) test whether bias increases the costs of integration. Columns (1) and (3) include school-by-cohort strata fixed effects, while columns (2) and (4) add a fixed effect for the ethnic pairing in the session (Macedonian-Albanian, Albanian-Turkish, Turkish-Macedonian). At the bottom of each column, we show the estimated θ, which theory predicts $\theta \leq 0$. Below it, we present the p-value associated with the one-sided test of the null hypothesis that $\theta > 0$.

	Network-level Learning			
	Hypothesis 2.1		Hypothesis 2.2	
	(1)	(2)	(3)	(4)
Correlated Signals	-0.030	-0.027		
	(0.02)	(0.02)		
Uncorrelated Priors X Integrated			-0.034	-0.034
			(0.03)	(0.03)
Correlated Priors X Segregated			-0.059	-0.056
			(0.04)	(0.04)
Correlated Priors X Integrated			-0.045	-0.040
			(0.03)	(0.03)
Constant	0.634***	0.632***	0.653***	0.651***
	(0.03)	(0.03)	(0.03)	(0.03)
Strata FE	Y	Y	Y	Y
Ethnic pairing FE	N	Y	N	Y
R-sqr	0.078	0.100	0.080	0.102
N	544	544	544	544
Theta	-0.030	-0.027	-0.048	-0.050
pval	0.106	0.139	0.173	0.154

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

Table A2: Robustness for Hypothesis 2, placebo. This table presents the regressions that test Hypothesis 2, only for the games in which names were hidden (anonymous play). In every column, the outcome is the network-level learning in the last round of each game, measured as the fraction of the 6 students that correctly guessed the color. Columns (1) and (2) test whether bias decreases learning. Columns (3) and (4) test whether bias increases the costs of integration. Columns (1) and (3) include school-by-cohort strata fixed effects, while columns (2) and (4) add a fixed effect for the ethnic pairing in the session (Macedonian-Albanian, Albanian-Turkish, Turkish-Macedonian). At the bottom of each column, we present the estimated θ which theory predicts $\theta \leq 0$. Below it, we present the p-value associated with the one-sided test of the null hypothesis that $\theta > 0$.

B.2 Robustness to round selection

	Network-level Learning			
	Hypothesis 1.1		Hypothesis 1.2	
	(1)	(2)	(3)	(4)
High Bias	-0.078*	-0.082*		
	(0.03)	(0.04)		
High Bias X Segregated			0.006	0.011
			(0.03)	(0.03)
Low Bias X Integrated			0.101*	0.109*
			(0.04)	(0.04)
Low Bias X Segregated			0.066	0.070
			(0.04)	(0.04)
Constant	0.686***	0.688***	0.605***	0.600***
	(0.06)	(0.06)	(0.06)	(0.06)
Strata FE	Y	Y	Y	Y
Ethnic pairing FE	N	Y	N	Y
R-sqr	0.120	0.136	0.122	0.140
N	2198	2198	2198	2198
Theta	-0.078	-0.082	-0.041	-0.050
pval	0.011*	0.012**	0.139	0.094*

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

Table A3: Robustness for Hypothesis 1, all rounds. This table presents the regressions that test Hypothesis 1, including all rounds of each game. In every column, the outcome is the network-level learning in the last round of each game, measured as the fraction of the 6 students that correctly guessed the color. Columns (1) and (2) test whether bias decreases learning. Columns (3) and (4) test whether bias increases the costs of integration. Columns (1) and (3) include school-by-cohort strata fixed effects, while columns (2) and (4) add a fixed effect for the ethnic pairing in the session (Macedonian-Albanian, Albanian-Turkish, Turkish-Macedonian). At the bottom of each column, we show the estimated θ, which theory predicts $\theta \leq 0$. Below it, we present the p-value associated with the one-sided test of the null hypothesis that $\theta > 0$.

	Network-level Learning			
	Hypothesis 2.1		Hypothesis 2.2	
	(1)	(2)	(3)	(4)
Correlated Signals	-0.023	-0.028		
	(0.02)	(0.02)		
Uncorrelated Priors X Integrated			-0.012	-0.009
			(0.02)	(0.02)
Correlated Priors X Segregated			-0.055*	-0.058*
			(0.03)	(0.03)
Correlated Priors X Integrated			0.007	0.004
			(0.03)	(0.03)
Constant	0.667***	0.669***	0.674***	0.675***
	(0.03)	(0.03)	(0.04)	(0.04)
Strata FE	Y	Y	Y	Y
Ethnic pairing FE	N	Y	N	Y
R-sqr	0.103	0.125	0.109	0.130
N	2198	2198	2198	2198
Theta	-0.023	-0.028	-0.074	-0.071
pval	0.155	0.112	0.019**	0.020**

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

Table A4: Robustness for Hypothesis 2, all rounds. This table presents the regressions that test Hypothesis 2, including all rounds of each game. In every column, the outcome is the network-level learning in the last round of each game, measured as the fraction of the 6 students that correctly guessed the color. Columns (1) and (2) test whether bias decreases learning. Columns (3) and (4) test whether bias increases the costs of integration. Columns (1) and (3) include school-by-cohort strata fixed effects, while columns (2) and (4) add a fixed effect for the ethnic pairing in the session (Macedonian-Albanian, Albanian-Turkish, Turkish-Macedonian). At the bottom of each column, we present the estimated θ which theory predicts $\theta \leq 0$. Below it, we present the p-value associated with the one-sided test of the null hypothesis that $\theta > 0$.

	Network-level Learning			
	Hypothesis 1.1		Hypothesis 1.2	
	(1)	(2)	(3)	(4)
High Bias	-0.087**	-0.082*		
	(0.03)	(0.04)		
High Bias X Segregated			0.017	0.020
			(0.03)	(0.03)
Low Bias X Integrated			0.103*	0.101*
			(0.04)	(0.04)
Low Bias X Segregated			0.093*	0.089*
			(0.04)	(0.04)
Constant	0.631***	0.629***	0.534***	0.536***
	(0.05)	(0.05)	(0.06)	(0.06)
Strata FE	Y	Y	Y	Y
Ethnic pairing FE	N	Y	N	Y
R-sqr	0.133	0.140	0.134	0.141
N	574	574	574	574
Theta	-0.087	-0.082	-0.027	-0.032
pval	0.004***	0.011**	0.233	0.189

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

Table A5: Robustness for Hypothesis 1, only first round. This table presents the regressions that test Hypothesis 1, including only the first round each game. In every column, the outcome is the network-level learning in the last round of each game, measured as the fraction of the 6 students that correctly guessed the color. Columns (1) and (2) test whether bias decreases learning. Columns (3) and (4) test whether bias increases the costs of integration. Columns (1) and (3) include school-by-cohort strata fixed effects, while columns (2) and (4) add a fixed effect for the ethnic pairing in the session (Macedonian-Albanian, Albanian-Turkish, Turkish-Macedonian). At the bottom of each column, we show the estimated θ, which theory predicts $\theta \leq 0$. Below it, we present the p-value associated with the one-sided test of the null hypothesis that $\theta > 0$.

	Network-level Learning			
	Hypothesis 2.1		Hypothesis 2.2	
	(1)	(2)	(3)	(4)
Correlated Signals	-0.029	-0.030		
	(0.02)	(0.02)		
Uncorrelated Priors X Integrated			-0.021	-0.021
			(0.02)	(0.02)
Correlated Priors X Segregated			-0.049	-0.052
			(0.03)	(0.03)
Correlated Priors X Integrated			-0.022	-0.022
			(0.03)	(0.03)
Constant	0.611***	0.612***	0.622***	0.623***
	(0.03)	(0.03)	(0.03)	(0.03)
Strata FE	Y	Y	Y	Y
Ethnic pairing FE	N	Y	N	Y
R-sqr	0.110	0.127	0.112	0.129
N	574	574	574	574
Theta	-0.029	-0.030	-0.049	-0.051
pval	0.111	0.100	0.073*	0.065*

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

Table A6: Robustness for Hypothesis 2, only first round. This table presents the regressions that test Hypothesis 2, including only the first round of each game. In every column, the outcome is the network-level learning in the last round of each game, measured as the fraction of the 6 students that correctly guessed the color. Columns (1) and (2) test whether bias decreases learning. Columns (3) and (4) test whether bias increases the costs of integration. Columns (1) and (3) include school-by-cohort strata fixed effects, while columns (2) and (4) add a fixed effect for the ethnic pairing in the session (Macedonian-Albanian, Albanian-Turkish, Turkish-Macedonian). At the bottom of each column, we present the estimated θ which theory predicts $\theta \leq 0$. Below it, we present the p-value associated with the one-sided test of the null hypothesis that $\theta > 0$.

B.3 Robustness to different bias proxies

	Network-level Learning			
	Hypothesis 1.1		Hypothesis 1.2	
	(1)	(2)	(3)	(4)
High Bias	-0.100*	-0.091*		
	(0.04)	(0.04)		
High Bias X Segregated			-0.013	-0.016
			(0.03)	(0.03)
Low Bias X Integrated			0.117*	0.105*
			(0.05)	(0.05)
Low Bias X Segregated			0.073	0.063
			(0.05)	(0.05)
Constant	0.711***	0.702***	0.618***	0.620***
	(0.05)	(0.05)	(0.02)	(0.02)
Strata FE	Y	Y	Y	Y
Ethnic pairing FE	N	Y	N	Y
R-sqr	0.117	0.137	0.120	0.140
N	574	574	574	574
Theta	-0.100	-0.091	-0.031	-0.026
pval	0.006***	0.019**	0.223	0.263

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

Table A7: Robustness for Hypothesis 1, language fluency as bias proxy. This table presents the regressions that test Hypothesis 1, but using fluency in local non-native languages as a low bias proxy. We calculate the reported fluency for each student, and then average this to create a group-level bias variable. The groups with below-median fluency are the *High Bias* groups, while those with above-median fluency are the *Low Bias* groups. In every column, the outcome is the network-level learning in the last round of each game, measured as the fraction of players who correctly guess the color in the last round. Columns (1) and (2) test whether bias decreases learning. Columns (3) and (4) test whether bias increases the costs of integration. Columns (1) and (3) include school-by-cohort strata fixed effects, while columns (2) and (4) add a fixed effect for the ethnic pairing in the session (Macedonian-Albanian, Albanian-Turkish, Turkish-Macedonian). At the bottom of each column, we show the estimated θ, which theory predicts $\theta \leq 0$. Below it, we present the p-value associated with the one-sided test of the null hypothesis that $\theta > 0$.

	Network-level Learning			
	Hypothesis 1.1		Hypothesis 1.2	
	(1)	(2)	(3)	(4)
High Bias	-0.072	-0.054		
	(0.04)	(0.04)		
High Bias X Segregated			-0.033	-0.033
			(0.03)	(0.03)
Low Bias X Integrated			0.073	0.055
			(0.05)	(0.05)
Low Bias X Segregated			0.040	0.023
			(0.04)	(0.04)
Constant	0.683***	0.665***	0.630***	0.630***
	(0.04)	(0.05)	(0.02)	(0.02)
Strata FE	Y	Y	Y	Y
Ethnic pairing FE	N	Y	N	Y
R-sqr	0.107	0.130	0.110	0.133
N	574	574	574	574
Theta	-0.072	-0.054	0.001	0.001
pval	0.026**	0.084*	0.512	0.513

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

Table A8: Robustness for Hypothesis 1, IOS as bias proxy. This table presents the regressions that test Hypothesis 1, but using Inclusion of Others in Self metric (Aron et al., 1992) as a low bias proxy. We calculate the reported proximity for each student toward the ethnic group they are playing with. Then, we average this to create a group-level bias variable. The groups with below-median proximity are the *High Bias* groups, while those with above-median proximity are the *Low Bias* groups. In every column, the outcome is the network-level learning in the last round of each game, measured as the fraction of players who correctly guess the color in the last round. Columns (1) and (2) test whether bias decreases learning. Columns (3) and (4) test whether bias increases the costs of integration. Columns (1) and (3) include school-by-cohort strata fixed effects, while columns (2) and (4) add a fixed effect for the ethnic pairing in the session (Macedonian-Albanian, Albanian-Turkish, Turkish-Macedonian). At the bottom of each column, we show the estimated θ, which theory predicts $\theta \leq 0$. Below it, we present the p-value associated with the one-sided test of the null hypothesis that $\theta > 0$.

	Network-level Learning			
	Hypothesis 1.1		Hypothesis 1.2	
	(1)	(2)	(3)	(4)
High Bias	-0.012	-0.026		
	(0.04)	(0.04)		
High Bias X Segregated			0.013	0.011
			(0.03)	(0.03)
Low Bias X Integrated			0.053	0.066
			(0.05)	(0.05)
Low Bias X Segregated			-0.005	0.008
			(0.05)	(0.04)
Constant	0.617***	0.624***	0.593***	0.587***
	(0.03)	(0.03)	(0.04)	(0.03)
Strata FE	Y	Y	Y	Y
Ethnic pairing FE	N	Y	N	Y
R-sqr	0.093	0.124	0.099	0.130
N	574	574	574	574
Theta	-0.012	-0.026	-0.070	-0.069
pval	0.383	0.258	0.040**	0.041**

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

Table A9: Robustness for Hypothesis 1, Dictator Game as bias proxy. This table presents the regressions that test Hypothesis 1, but using a Dictator Game fairness as a low bias proxy. We define Dictator Game fairness as 1 if students split money equally or gave more to other ethnicities than their own. We then average it to create a group-level bias variable. The groups with below-median dictator game bias are the *High Bias* groups, while those with above-median proximity are the *Low Bias* groups. In every column, the outcome is the network-level learning in the last round of each game, measured as the fraction of players who correctly guess the color in the last round. Columns (1) and (2) test whether bias decreases learning. Columns (3) and (4) test whether bias increases the costs of integration. Columns (1) and (3) include school-by-cohort strata fixed effects, while columns (2) and (4) add a fixed effect for the ethnic pairing in the session (Macedonian-Albanian, Albanian-Turkish, Turkish-Macedonian). At the bottom of each column, we show the estimated θ, which theory predicts $\theta \leq 0$. Below it, we present the p-value associated with the one-sided test of the null hypothesis that $\theta > 0$.

	Network-level Learning			
	Hypothesis 1.1		Hypothesis 1.2	
	(1)	(2)	(3)	(4)
High Bias	-0.048	-0.055		
	(0.04)	(0.04)		
High Bias X Segregated			0.002	0.008
			(0.04)	(0.04)
Low Bias X Integrated			0.089	0.104
			(0.05)	(0.05)
Low Bias X Segregated			0.020	0.025
			(0.05)	(0.05)
Constant	0.627***	0.630***	0.578***	0.571***
	(0.04)	(0.04)	(0.05)	(0.05)
Strata FE	Y	Y	Y	Y
Ethnic pairing FE	N	Y	N	Y
R-sqr	0.115	0.134	0.120	0.142
N	409	409	409	409
Theta	-0.048	-0.055	-0.070	-0.087
pval	0.127	0.101	0.068*	0.035**

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

Table A10: Robustness for Hypothesis 1, 1^{st} and 2^{rd} terciles as bias proxy. This table presents the regressions that test Hypothesis 1, but using a tercile split on the *Any cross-ethnic friend* group-level variable. The tercile is 1 if the group has between 0 and 2 players with a cross-ethnic friend, 2 if the group has exactly 3 players with a cross-ethnic friend, and 3 if the group has between 4 and 6 players with a cross-ethnic friend. In this table we code the first tercile as *High Bias* groups and the second tercile as *Low Bias* groups. In this table, the groups in the first tercile are the *High Bias* groups, while those in the second tercile are the *Low Bias* groups. In every column, the outcome is the network-level learning in the last round of each game, measured as the fraction of players who correctly guess the color in the last round. Columns (1) and (2) test whether bias decreases learning. Columns (3) and (4) test whether bias increases the costs of integration. Columns (1) and (3) include school-by-cohort strata fixed effects, while columns (2) and (4) add a fixed effect for the ethnic pairing in the session (Macedonian-Albanian, Albanian-Turkish, Turkish-Macedonian). At the bottom of each column, we show the estimated θ, which theory predicts $\theta \leq 0$. Below it, we present the p-value associated with the one-sided test of the null hypothesis that $\theta > 0$.

	Network-level Learning			
	Hypothesis 1.1		Hypothesis 1.2	
	(1)	(2)	(3)	(4)
High Bias	-0.133**	-0.124**		
	(0.04)	(0.04)		
High Bias X Segregated			0.011	0.015
			(0.04)	(0.04)
Low Bias X Integrated			0.164**	0.157**
			(0.05)	(0.05)
Low Bias X Segregated			0.121*	0.114*
			(0.06)	(0.05)
Constant	0.799***	0.791***	0.661***	0.659***
	(0.04)	(0.04)	(0.02)	(0.02)
Strata FE	Y	Y	Y	Y
Ethnic pairing FE	N	Y	N	Y
R-sqr	0.174	0.180	0.177	0.182
N	404	404	404	404
Theta	-0.133	-0.124	-0.054	-0.057
pval	0.002***	0.002***	0.140	0.128

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

Table A11: Robustness for Hypothesis 1, 1^{st} and 3^{rd} terciles as bias proxy. This table presents the regressions that test Hypothesis 1, but using a tercile split on the *Any cross-ethnic friend* group-level variable. The tercile is 1 if the group has between 0 and 2 players with a cross-ethnic friend, 2 if the group has exactly 3 players with a cross-ethnic friend, and 3 if the group has between 4 and 6 players with a cross-ethnic friend. In this table we code the first tercile as *High Bias* groups and the second tercile as *Low Bias* groups. In this table, the groups in the first tercile are the *High Bias* groups, while those in the third tercile are the *Low Bias* groups. In every column, the outcome is the network-level learning in the last round of each game, measured as the fraction of players who correctly guess the color in the last round. Columns (1) and (2) test whether bias decreases learning. Columns (3) and (4) test whether bias increases the costs of integration. Columns (1) and (3) include school-by-cohort strata fixed effects, while columns (2) and (4) add a fixed effect for the ethnic pairing in the session (Macedonian-Albanian, Albanian-Turkish, Turkish-Macedonian). At the bottom of each column, we show the estimated θ, which theory predicts $\theta \leq 0$. Below it, we present the p-value associated with the one-sided test of the null hypothesis that $\theta > 0$.

B.4 Robustness to differences due to network structure

	Network-level Learning			
	Hypothesis 1.1		Hypothesis 1.2	
	(1)	(2)	(3)	(4)
High Bias	-0.093*	-0.092*		
	(0.04)	(0.04)		
High Bias X Segregated			0.052	0.058
			(0.04)	(0.04)
Low Bias X Integrated			0.129**	0.134**
			(0.05)	(0.05)
Low Bias X Segregated			0.120*	0.121*
			(0.05)	(0.05)
Constant	-0.177***	-0.178***	-0.300***	-0.302***
	(0.05)	(0.05)	(0.06)	(0.06)
Strata FE	Y	Y	Y	Y
Ethnic pairing FE	N	Y	N	Y
R-sqr	0.116	0.140	0.120	0.144
N	574	574	574	574
Theta	-0.093	-0.092	-0.060	-0.071
pval	0.005***	0.009***	0.079*	0.048**

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

Table A12: Robustness for Hypothesis 1, normalized learning. This table presents the regressions that test Hypothesis 1. In every column, the outcome is the network-level normalized learning in the last round of each game, measured as the difference between model-implied learning under no bias and actual learning. Columns (1) and (2) test whether bias decreases learning. Columns (3) and (4) test whether bias increases the costs of integration. Columns (1) and (3) include school-by-cohort strata fixed effects, while columns (2) and (4) add a fixed effect for the ethnic pairing in the session (Macedonian-Albanian, Albanian-Turkish, Turkish-Macedonian). At the bottom of each column, we show the estimated θ, which theory predicts $\theta \leq 0$. Below it, we present the p-value associated with the one-sided test of the null hypothesis that $\theta > 0$.

	Network-level Learning			
	Hypothesis 2.1		Hypothesis 2.2	
	(1)	(2)	(3)	(4)
Correlated Signals	-0.017	-0.022		
	(0.02)	(0.02)		
Uncorrelated Priors X Integrated			-0.040	-0.038
			(0.02)	(0.02)
Correlated Priors X Segregated			-0.043	-0.047
			(0.03)	(0.03)
Correlated Priors X Integrated			-0.023	-0.027
			(0.04)	(0.04)
Constant	-0.203***	-0.201***	-0.183***	-0.182***
	(0.02)	(0.02)	(0.03)	(0.03)
Strata FE	Y	Y	Y	Y
Ethnic pairing FE	N	Y	N	Y
R-sqr	0.094	0.125	0.097	0.128
N	574	574	574	574
Theta	-0.017	-0.022	-0.060	-0.058
pval	0.242	0.185	0.053*	0.058*

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

Table A13: Robustness for Hypothesis 2, normalized learning. This table presents the regressions that test Hypothesis 2. In every column, the outcome is the network-level normalized learning in the last round of each game, measured as the difference between model-implied learning under no bias and actual learning. Columns (1) and (2) test whether bias decreases learning. Columns (3) and (4) test whether bias increases the costs of integration. Columns (1) and (3) include school-by-cohort strata fixed effects, while columns (2) and (4) add a fixed effect for the ethnic pairing in the session (Macedonian-Albanian, Albanian-Turkish, Turkish-Macedonian). At the bottom of each column, we present the estimated θ which theory predicts $\theta \leq 0$. Below it, we present the p-value associated with the one-sided test of the null hypothesis that $\theta > 0$.