# Diversity and Cooperation

Dan Anderberg* Gordon B. Dahl† Christina Felfe‡ Helmut Rainer§ Thomas Siedler¶

October 20, 2024

**Abstract:** What makes diversity unifying in some settings but divisive in others? We examine how the mixing of ethnic groups in German schools affects intergroup cooperation and trust. We leverage the quasi-random assignment of students to classrooms within schools to obtain variation in the type of diversity that prevails in a peer group. We combine this with a large-scale, incentivized lab-in-field experiment based on the investment game, allowing us to assess the in-group bias of native German students in interactions with other natives (in-group) versus immigrants (out-group). We find in-group bias peaks in culturally polarized classrooms, where the native and immigrant groups are both large, but have different religious backgrounds. In contrast, in classrooms characterized by non-cultural polarization, fractionalization, or a native supermajority, there are significantly lower levels of own-group favoritism. Exploring mechanisms, we find that in culturally polarized classrooms, natives are less likely to form friendships with immigrants, have less of a shared identity, and experience less favorable treatment from immigrants than from fellow natives. These classroom experiences foster negative stereotypes, reducing trust and diminishing other-regarding preferences toward immigrants outside the classroom. Our findings suggest that extra efforts are needed to counteract low levels of inclusivity and trust in culturally polarized environments.

**Keywords:** In-group bias, diversity, cultural polarization
**JEL codes:** J15

**Acknowledgments:** We thank Samuel Bazzi, V. Bhaskar, Paul Niehaus, and David Yanagizawa-Drott, as well as seminar participants at several universities and conferences. We are grateful to the almost 20 research assistants and interns for their invaluable assistance with data collection. The project received generous financial support from the ifo Institute, University of Munich, University of St. Gallen, and University of Hamburg.

*Royal Holloway University of London, CESifo, IFS (email: dan.anderberg@rhul.ac.uk)

†UC San Diego, Norwegian School of Economics, NBER, ifo Institute, CESifo, CEPR, CReAM, IZA (email: gdahl@ucsd.edu)

‡University of Konstanz, CESifo, CEPR, IZA (email: christina.felfe@uni-konstanz.de)

§University of Munich, ifo Institute, CESifo (email: rainer@econ.lmu.de)

¶University of Potsdam (email: thomas.siedler@uni-potsdam)

# 1 Introduction

Immigration has surged in recent decades, rendering societies increasingly diverse. In North America and Europe, the number of international migrants has risen by roughly 50% in the past 20 years, leading to an influx of individuals with culturally diverse backgrounds (McAuliffe and Triandafyllidou, 2021). Forecasts indicate a continued rise in immigration in the years to come (Hanson and McIntosh, 2016), which will perpetuate ongoing societal transformations. Sustaining social cohesion amid this heightened diversity requires trust and cooperation across national, religious, and ethnic divides (Algan, 2018).

Public education stands out as one of the few social institutions with the potential to foster such intergroup cooperation. Indeed, one argument for publicly provided schooling is to create a unified citizenry (Dee, 2004; Milligan et al., 2004). Schools could play a crucial role in promoting greater cross-cultural understanding by offering opportunities for youth from diverse backgrounds to form meaningful relationships across group boundaries. This, in turn, could break down prejudices and stereotypes, ultimately preparing them to thrive in and contribute positively to a diverse society (Allport, 1954; Pettigrew and Tropp, 2008; Wells et al., 2016; Tropp and Saxena, 2018).

As intuitively appealing as this view might be, it ignores that diversity may facilitate *or* obstruct intergroup cooperation depending on the form it takes. In settings dominated by a few large groups (polarization), the drive to establish cultural domination could grow strong, causing own-group attachment to increase and integration to decrease. In contrast, in environments where several small groups coexist (fractionalization), substantial benefits could be gained by unifying under a shared identity, which in turn would ultimately improve social cohesion.

The opposing forces of polarization and fractionalization have been explored in the context of civil conflict and nation-building in the developing world (e.g., Montalvo and Reynal-Querol, 2005; Bazzi et al., 2019). Our study brings the fractionalization-polarization paradigm to bear on the issue of intergroup cooperation in diverse classrooms in Germany. Our interest lies in whether the type of diversity that prevails in schools matters for in-group bias in trust and cooperation. Specifically, are adolescents embedded in polarized peer groups more likely to display own-group favoritism compared to adolescents with highly homogeneous or fractionalized peer groups? And what role does the cultural distance between majority and minority groups play in this?

Our study is unique for three reasons. First, we examine these forces during adolescence, a critical life-cycle stage where an individual's attitudes and biases are susceptible to socializing influences. Schools are a setting where youth interact with peers from diverse backgrounds on a daily basis, and hence create an environment where both intergroup conflict and cooperation could emerge (Allport, 1954; Depetris-Chauvin et al., 2020; Lowe, 2021; Mousa, 2020). Second, we zoom in on the micro level of classrooms, whereas most of the existing literature on intergroup cooperation has focused on nations and communities. Third, we study a highly developed country where immigration is the driving force for changes in diversity, a trend which will only accelerate across Europe due to demographic and economic pressures.

Estimating the impacts of polarization and fractionalization is challenging for two reasons. First, data on in-group out-group cooperation among adolescents is scarce, and even when available, cannot readily be linked to individuals' peer groups. Second, the set of adolescents who interact with ethnically diverse peers is likely to be endogenous, resulting in selection bias. For example, youth embedded in ethnically diverse peer groups may be more open to diversity to begin with, while those exposed to an ethnically homogeneous learning environment may have parents who transmit prejudicial attitudes.

Our paper overcomes these challenges by using a unique design that combines two elements. First, on the data front, we run a large, incentivized lab-in-the field experiment in 220 classes spread across 57 German secondary schools. The experiment, based on an investment (or "trust") game, allows us to measure how native German students cooperate with in-group (other natives) versus out-group (immigrants) partners, respectively. In the game, the sender chooses how much money to transfer, this amount is multiplied by 3, and the receiver decides how much money to transfer back. Key to our experiment is that each participant is asked how they will play if paired with a native and an immigrant interaction partner, both of whom are anonymous to the participant as they are drawn from a different school. Our main measure for in-group bias in cooperation is how much a sender chooses to transfer if they are paired with a native versus an immigrant.[1]

In addition, we ask students to fill out an extensive survey, allowing us to characterize classrooms in terms of their ethnic composition. Forty-six percent of students in our sample have an immigrant background, spanning in total 95 different countries of origin, with the largest groups coming from Turkey, Poland, and Russia. Indeed, Germany has the second largest number of international migrants behind the U.S. (McAuliffe and Triandafyllidou, 2021), providing a rich testing ground to study the effects of ethnic and cultural diversity.

Second, in terms of identification, we exploit variation in peer group diversity arising from students' quasi-random assignment to classes within schools. As we explain later, in Germany, all students have the right to receive a non-discriminatory education, and schools often have explicit rules to ensure this applies in the assignment of students to classrooms. Consistent with this, we conduct several empirical tests which provide strong support that native and immigrant students are randomly assigned to classrooms within a school. We first verify that, once school fixed effects are accounted for, individual and family background characteristics are not statistically significant predictors of classroom diversity. We also document that the actual distribution of immigrant peers closely matches the simulated distribution assuming random assignment.

Combining this variation in classroom diversity with our experimentally elicited data, we first explore how natives' in-group bias varies with the fraction of immigrants in a class, regardless of the immigrants' origin. We estimate an inverse-U shaped relationship: the in-group bias among

[1]Participants are asked to play the game as both sender and receiver. We use the decisions of receivers to study mechanisms.

natives initially widens as the fraction of immigrants increases, but eventually turns around and narrows again as the immigrant fraction continues to increase. The turning point occurs where there is close to an even split of natives versus immigrants in a class. In other words, when natives are the dominant majority, adding in extra immigrants to a classroom increases the in-group bias among natives. But once natives start to lose their majority status, adding in extra immigrants decreases the bias. Around the turning point, a one-standard deviation change in the immigrant share (21 pp) shrinks natives' in-group bias by 10% of a standard deviation.

In a second step, we investigate how the cultural background and relative size of immigrant subgroups matter. To that end, we no longer treat immigrants in a classroom as a single group, but consider cultural heterogeneity among them. We use religious affiliation as the differentiating factor, specifically distinguishing between immigrants who are Muslim and those who are not. This split allows us to compare immigrants who are culturally distant from their native peers versus those who are culturally more similar. We estimate that the highest level of in-group bias among natives occurs in classrooms characterized by cultural polarization, where a slight majority of natives coexists with a substantial minority of Muslim immigrants.

In sharp contrast, in classrooms exhibiting other forms of diversity, natives' tendency towards in-group favoritism is 32% to 42% of a standard deviation lower compared to culturally polarized classrooms. This encompasses scenarios with (i) *non-cultural polarization*, where a slight majority of natives coexist with a substantial minority of non-Muslim immigrants, (ii) *mixed polarization*, where there is a slight majority of natives and immigrants are equally divided between Muslims and non-Muslims, (iii) *overall fractionalization*, where natives, Muslim immigrants, and non-Muslim immigrants are equally large groups, and (iv) a *native supermajority*. The pronounced in-group bias in culturally polarized classrooms carries significant costs. We estimate that the limited out-group cooperation in culturally-polarized environments results in payoff losses which are between 0.30 to 0.44 standard deviations higher compared to classrooms with non-cultural polarization, mixed polarization, overall fractionalization, or a native supermajority.

We next explore potential mechanisms. In our experiment, participants are paired with both native and immigrant interaction partners drawn from the general student population (i.e., not their own classroom or school). As summarized above, natives in culturally polarized classrooms display pronounced in-group biases, leading to financial losses. Our main hypothesis to explain this behavior consists of two parts. First, natives in culturally polarized classrooms may develop mistrust and negative sentiments toward immigrants because of limited or adversarial interactions with them in their classroom. Second, if natives generalize inaccurately from these experiences, they may form negative stereotypes, reducing trust and other-regarding preferences toward immigrants more broadly. Given that both trust and other-regarding preferences are important rationales for transferring money in the investment game (see, e.g., Sapienza et al., 2013), this could explain the heightened in-group bias observed within culturally polarized classrooms versus other classroom scenarios.

Three empirical findings support the first part of this hypothesis. First, culturally polarized classrooms inhibit the formation of friendships between natives and immigrants. This conclusion is drawn using a survey question that enables us to create an intergroup friendship index. We estimate the index to be between 37% and 96% of a standard deviation lower in culturally polarized settings compared to non-culturally polarized, mixed polarized, overall fractionalized, and native supermajority classrooms. Second, using a survey question that measures the difference in how strongly natives and immigrants self-identify as German, we find that natives and immigrants in culturally polarized classrooms are much less likely to develop shared identities than in other classroom contexts. The identity gap between natives and immigrants is 56% to 94% of standard deviation higher in culturally polarized classrooms than in the four other scenarios. Third, within culturally polarized classrooms, natives are treated less favorably by immigrants compared to native peers, a result based on a proxy measure for reciprocity among classroom peers. Therefore, while a culturally polarized classroom provides considerable exposure to immigrants, this type of classroom appears to hinder positive interactions between natives and immigrants.

The second part of our hypothesis posits that these classroom experiences lead natives to develop negative stereotypes, resulting in reduced trust and lower other-regarding preferences toward immigrants in general. This is supported by three empirical findings. First, we use questions from our survey which capture respondents' trust towards natives and immigrants in the general population. We estimate that the in-group out-group trust gap among natives reaches its peak in culturally polarized classrooms. The gap is between 33% to 50% of a standard deviation higher when contrasted with non-culturally polarized, mixed polarized, overall fractionalized, and native majority classrooms. Second, we examine other-regarding preferences—i.e., the idea that natives simply prefer to give to fellow natives over immigrants—by leveraging the strategy-vector method in our experiment. This approach makes the choices of students when playing the role of receiver in our investment game akin to the choices made in a standard dictatorship game. We find that preferences for fellow natives over immigrants is more pronounced in culturally polarized classrooms than in other classroom settings. Third, we use a survey-based measure of anti-immigrant sentiment related to the labor market as a proxy for other-regarding preferences. This measure of anti-immigrant sentiment also peaks in culturally polarized classrooms, with an estimate which is significantly different from the other four classroom types.

To help interpret our findings, we outline a simple model which integrates a social interactions model à la Brock and Durlauf (2001) into a Hotelling-type structure (Hotelling, 1929) with multiple interacting types. In a version of the model tailored to our empirical setting, individuals fall into three types: natives, culturally close immigrants, and culturally distant immigrants. They face the choice of *either* joining an "inclusive" group with a common shared identity *or* preserving their innate cultural identity in an "exclusive" group composed solely of others of their own type. This decision entails a trade-off between the desire to be part of a large group and a preference for one's own cultural identity. The identity of the inclusive group shifts endogenously with the distribution of types in the population. Due to a social multiplier, small shifts in the population type distribution can lead to large changes in inclusivity. Specifically, consistent with our empirical evidence, comparative statics show a low willingness of natives to engage with immigrants in the inclusive group under cultural polarization, as opposed to scenarios involving non-cultural polarization, mixed polarization, fractionalization, or a native supermajority.

Our study contributes to several strands of research. Diversity and its consequences for social cohesion have long been studied in economics. Examples of earlier work include Alesina and La Ferrara (2000, 2002) and Luttmer (2001), while Alesina and Tabellini (2024) provide an up-to-date survey. Recent evidence indicates that diversity can either hinder or improve social cohesion depending on the context. In France, it has been shown to strain social bonds among neighbors and reduce the quality of local public goods (Algan et al., 2016). However, in the U.S., the outmigration of millions of African Americans from the South to the North in the mid-20th century was associated with increased support among whites for racial equality (Calderon et al., 2023). A study in Danish classrooms by Kruse (2023) found that rising diversity bolsters natives' solidarity with immigrants. Using data from Russia and the U.S., Egorov et al. (2021) showed that prosocial behavior increased in more ethnically diverse communities during the COVID-19 pandemic. Our work diverges from these studies by focusing on in-group biases in trust and cooperation and separating the effects of diversity into its polarization and fractionalization components.

To our knowledge, only one previous study has used individual-level data to disentangle these effects. Bazzi et al. (2019) examine a resettlement program in Indonesia which relocated millions of ethnically diverse migrants. In cases where the program led to fractionalized communities, individuals developed shared identities, whereas in polarized communities, attachment to one's own ethnic group remained strong. Like Bazzi et al. (2019), we hone in on the different components of diversity, but study its consequences in a wealthy country where migration-induced diversity is only going to become more pervasive in the years ahead. In contrast to a long-standing argument that diversity is less divisive in rich nations (Horowitz, 1985), we find that polarization and fractionization matter for social cohesion in a way akin to the developing world.

Our study also relates to recent work examining social cohesion in school and university settings. For example, Alan et al. (2021) study a perspective-taking intervention in the aftermath of the influx of Syrian refugees into Turkish elementary schools, showing that it contributed to fewer physical conflicts among peers, less social exclusion, and improved inter-ethnic social ties. Boucher et al. (2022) examine a classroom intervention where five-year old Turkish and Syrian refugee children were randomly brought into contact, discovering that such exposure resulted in an increased formation of interethnic friendships. Rao (2019) exploits a policy change in India, where poor students were integrated into elite private schools, finding that economically disadvantaged classmates makes wealthier students more prosocial, generous, and egalitarian. Carrell et al. (2019) and Corno et al. (2022) leverage the random assignment of roommates in university settings in the U.S. and South Africa to demonstrate that students become more empathetic and open to forming friendships with members of the social groups to which their roommates belong. These studies suggest that interventions can improve social cohesion and that intergroup contact promotes integration. In contrast, our study highlights that classroom contact alone might not be sufficient, but in fact can be counteracted by own-group attachment in culturally polarized classrooms.

As emphasized in the literature, peer groups are central in the socialization of adolescents, playing a key role in shaping their attitudes and behaviors (Brown, 2011). A large literature studies peer effects for outcomes as diverse as education, crime, drug use, and teenage pregnancy (see, e.g., Kremer and Levy, 2008; Bifulco et al., 2011; Sacerdote, 2011; Lavy et al., 2012; Ohinata and Van Ours, 2013; Brenøe and Zölitz, 2020; Chuard et al., 2022; Figlio et al., 2024). Our paper adds to this literature by studying how the type of diversity in a peer group matters (polarization versus fractionalization) and by examining intergroup cooperation and trust.

In our own previous work (Felfe et al., 2021), we utilized the experimental data we collected from German schools to investigate a completely different question for a different group. Specifically, we studied how immigrants' prosocial behavior was affected by a reform which granted them citizenship from birth. We also used the survey data we collected to show that the same birthright citizenship reform had unintended consequences for the well-being of immigrant girls (Dahl et al., 2022).

The remainder of the paper is organized as follows. We first explain our setting, the experiment, and the survey. In Section 3, we discuss our empirical design and how we model classroom diversity. Section 4 presents our results on how in-group bias varies both with the amount and type of diversity in a classroom, and Section 5 explores possible mechanisms. The following section outlines a simple model which is consistent with our empirical findings. The final section concludes.

# 2 Study Design

## 2.1 Setting

We study how diversity affects in-group bias in the context of classrooms in Germany, which have varying mixes of native and immigrant students. We received permission to enter classrooms and collect data from all students completing their final year of secondary school for 57 schools (222 classrooms). These students were either in $9^{th}$ or $10^{th}$ grade, and hence mostly 15 or 16 years old. In the German school system, when students enter secondary school in fifth grade, they are assigned to a class. Typically these students remain together in the same class until the end of compulsory school, making their exposure to classroom peers long term.

We collected data in two distinct phases. From June 2 and July 15, 2015, we visited 31 schools across five cities in the German state of Schleswig Holstein (SH), where we gathered data from all 122 ninth-grade classes. From October 19 to November 16, we surveyed students in 100 tenth-grade classes located in 26 schools in two cities within the state of North Rhine Westphalia (NRW).[2]

[2]In both federal states, a school year starts in August/September and ends in June/July. One important difference is that compulsory schooling lasts 9 years in SH, but 10 years in NRW. Thus, in our sample, students from the

Our data collection consisted of two main components. First, we implemented an incentivized lab-in-the-field experiment to measure native and immigrant students' willingness to cooperate with in-group versus out-group partners. Second, we administered a survey to collect comprehensive family background information, including details on the students' ethnicity and religious affiliation.

The research was conducted during regular class periods, when students would otherwise be engaged in standard classroom learning. The regular teacher either remained inconspicuous at the back of the classroom or left the room. Meanwhile, members of the research team and/or trained research assistants introduced and supervised the study. Students were seated at their usual desks, with mobile privacy screens placed between them. The study was administered using traditional paper and pencil methods and spanned two consecutive class periods, each lasting 45 minutes. Whether the experiment or the survey was administered first was randomized on a daily basis.

Our study was planned and executed in close collaboration with the education ministries of the federal states of Schleswig-Holstein and North Rhine Westphalia. Prior to commencing our research, we sought approval from these ministries and obtained consent from school principals. These school principals, in turn, communicated with parents to inform them about the upcoming study and allowed them to opt out. Out of the 4,634 students present at school on the day of the study, only 44 parents (<1%) chose not to have their child participate. At the beginning of each session, we emphasized to the students that their participation was entirely voluntary and anonymous. To ensure anonymity, every student received a unique, anonymous identity. 154 students (3.5%) chose not to take part in our study, leaving us with 4,436 students. Of those, 222 did not provide basic survey information essential for our analysis, namely, the respondent's gender, whether their parents were born in Germany, their parents' countries of birth, and the respondent's religious affiliation. Thus, our baseline sample comprises 4,214 students, distributed across 222 classes in 57 schools.

## 2.2 Measuring Classroom Diversity

To measure classroom diversity, we first use the survey data we collected to categorize students as having an immigrant background if at least one parent was born abroad or at least one parent has not obtained German citizenship. We further classify Muslim versus non-Muslim immigrants based on whether the student self-identifies as Muslim.[3] In our baseline sample of 4,214 students, 2,222 (53%) are natives, while 1,992 (47%) are immigrants.[4]

two states are from the same school starting cohort. Our secondary schools span various types: 10 general schools ("Hauptschule"); 8 intermediate schools ("Realschule"); 29 comprehensive schools without the final years of grammar school education ("Gesamtschule ohne gymnasiale Oberstufe"); 8 comprehensive schools with the final years of grammar school education ("Gesamtschule mit gymnasialer Oberstufe"); and 2 grammar schools or high schools ("Gymnasium").

[3]There are 29 students who self-identify as Muslim but who we do not classify as having an immigrant background. If we instead classify these 29 observations as having an Muslim immigrant background, if anything, the results become slightly stronger.

[4]In the general population, the share of individuals aged 15-20 with an immigrant background was 34% as of 2018 (Federal Statistical Office of Germany, 2009); our share is higher because we purposefully targeted cities with many immigrants.

The main goal of our analysis is to examine how the in-group bias of native German adolescents are affected by the diversity of their classroom peer groups, and the role played by cultural distance. We explore two dimensions of peer group diversity: (i) native-immigrant diversity, which is based on the share of immigrants in a classroom, irrespective of cultural background; and (ii) within-immigrant diversity, stemming from the cultural heterogeneity and relative sizes of immigrant subgroups within a classroom.

To measure native-immigrant diversity, we compute the proportion of classmates with an immigrant background as the leave-one-out share.[5] On average, native students are in classrooms where 39% of their peers are immigrants, with a standard deviation of 22% (see Table 1). Figure 1 shows two histograms: one for the number of classes with different shares of immigrant peers (left panel) and one for the number of native student observations in classes with different shares of immigrant peers (right panel). As we will show in the next section, even after we take out school fixed effects, there is wide variation in the fraction of immigrants in a class.

To study the role of within-immigrant diversity, instead of treating immigrants in a classroom as a homogeneous group, we consider their cultural heterogeneity. For our main set of results, we focus on a binary categorization which separates immigrants into those who are culturally close to versus far from their native peers. In the context of Germany, a natural criterion for assessing cultural distance is religion. Specifically, we differentiate between immigrants with a Muslim background and those with a non-Muslim background. Table 1 reveals that, on average, 19% of a native's classmates have a Muslim immigrant background (std. dev.=18%) and 20% have a non-Muslim immigrant background (std. dev.=11%). As we will show in the next section, even after we take out school fixed effects, there is wide variation in the fraction of Muslim versus non-Muslim immigrant peers in a classroom. In a robustness check, we also explore the impact of classroom polarization versus fractionalization by decomposing classrooms into a native group and 11 distinct immigrant subgroups based on their parents' country or region of origin.

### 2.3 Measuring In-Group Bias

**The Experiment.** To measure in-group bias, we used a modified version of the investment game originally introduced by Berg (1995). This choice was informed by the expanding body of research on discrimination between real social groups, which can be traced back to Fershtman and Gneezy's (2001) seminal work.

The investment game is a two-player scenario, with one participant acting as the sender (first-mover) and the other as the receiver (second-mover). Both players start with an initial endowment, in our

[5] To compute this share and other measures of classroom diversity, we use survey information from all of the 4,214 students in our baseline sample, of which 2,228 are natives. 54 native students failed to complete the first stage of our experiment. Two classes of the 222 classes in the baseline sample only have immigrants, and hence are not used in our analysis. Therefore, our estimation sample contains 2,168 native students spread over 220 classes. Summary statistics correspond to the 220 classrooms attended by these 2,168 native students.

case, 5 euros. The sender makes the initial decision regarding how much of their endowment to send to the receiver, $x \in [0, 5]$, with the constraint that they can only send in 50 cent increments. We then triple the amount sent to the receiver. The receiver, now possessing an amount of $5 + 3x$ euros, decides how much to send back to the sender. They can send an amount $y \in [0, 5 + 3x]$ in 10 cent increments. Consequently, the sender exits the game with $5 - x + y$ euros, while the receiver exits with $5 + 3x - y$ euros.[6]

In our experiment, we utilized the strategy method, which meant that each participant had to make decisions both as the first-mover and as the second-mover. We first had participants assume the role of first-movers. Importantly, first-movers had to factor in the gender and migration background of their potential interaction partners when making investment decisions. This was achieved by letting first-movers decide the amounts they wished to allocate to a male with German parents ($I_1$), a female with German parents ($I_2$), a male with foreign parents ($I_3$), and female with foreign parents ($I_4$).[7]

After participants completed the initial stage of the investment game, they were asked to specify their expected back transfer from each of the four possible interaction partners. This expectation was recorded within a range of 0 to 20 euros, in increments of ten cents.

In the final stage of the investment game, participants assumed the role of second-movers, and we employed the contingent response method to elicit their back transfers (returns). For instance, on one decision sheet, participants were tasked with determining their back transfers to a male with German-born parents for each of the eleven possible investments (0, .5, 1, 1.5, ..., 5) made by a male with German-born parents as the first mover. Using a similar approach, we collected back payments for the other potential interaction partners. Participants were allowed to specify amounts between 0 and $5 + 3x$ euros, in ten cent increments.

Prior to commencing the experiment, written instructions were provided to all students, with an experimenter verbally going through the instructions with the students as well. The translated instructions and the decision sheets are available in Appendix C. The students were informed that they would engage in the investment game, initially as the first-mover and subsequently as the second-mover. Students were told they had the opportunity to earn actual money, and their eventual earnings would be contingent on their individual decisions as well as those of another participant. Importantly, we explicitly told students they would be randomly assigned to play with an anonymous student from a completely different school (i.e., not from their own class or school).

To be more concrete, we informed participants that we would determine their ultimate earnings through the following steps: (i) we would randomly pair two participants from different schools in

[6] Assuming self-interested preferences, the only subgame-perfect equilibrium has no investment and zero returns. In contrast, "full" cooperation, where the first-mover invests their entire endowment, maximizes the players' joint payoff.

[7] Additionally, we inquired about the amounts students would be willing to send to a boy with foreign parents holding a German passport ($I_5$) and a girl with foreign parents who possessed a German passport ($I_6$). This line of questioning was included in the data collection process because our study was initially designed to explore the integration of immigrant children who had acquired German citizenship.

the same federal state; (ii) we would then randomly assign the roles of first-mover and second-mover; (iii) the relevant characteristics of both the first-mover and the second-mover, namely gender and immigration background, would be established using survey information; (iv) we would execute the investment decision made by the first-movers using the characteristics of the second-mover; (v) the back transfer decision of the second-mover would be implemented, taking into account the characteristics of the first-mover and their investment choice from step (iv); (vi) finally, based on the combination of choices from steps (iv) and (v), we would calculate the participants' ultimate earnings. Payments were disbursed within a two-week timeframe and delivered in sealed envelopes, each bearing the student's distinctive identification code, by either the head teacher or the school's secretary.

In applying this method to determine participants' earnings, we categorized mixed-background children (those with one German-born and one foreign-born parent) as having foreign-born parents.[8] This categorization is consistent with how we define our measures of classroom diversity in Section 2.2.

**In-Group Out-Group Investment Gap.** To create our main dependent variable, we rely on the investment decisions of native students during the first stage of the investment game. On average, native students invest 2.86 euros, which corresponds to 57 percent of their initial 5 euro endowment. Figure 2 shows a histogram of all possible investment decisions made by the native students in our sample. The two most frequent investment choices are transfers equivalent to either 50% or 100% of their initial endowment. These patterns resemble those found in similar lab-in-the-field experiments.[9]

To measure the in-group bias in cooperation among native students, we take the difference between their average investments in fellow natives averaged over both genders, given by $\frac{1}{2}(I_1 + I_2)$, and their average investments in immigrants, given by $\frac{1}{2}(I_3 + I_4)$. We denote this measure as the in-group out-group investment gap (IG). The summary statistics for this measure are included in Table 1. Among native students, the average IG is 0.09 euros and the standard deviation is 0.76. Hence, on average there is not much in-group favoritism, but it varies considerably.

We note that experimenter demand effects and social desirability bias should be constant across classroom types, unless these factors are directly affected by diversity. If they are directly affected by classroom diversity, then this is a mechanism for our findings, rather than a threat to identification.

## 2.4 Estimation Sample

Our main estimation sample consists of 2,168 native German students. Summary statistics in Table 1 indicate that these students were 15.8 years old on average at the time of the study. Approximately

[8] There were no questions from participants about the handling of mixed-background children during the experiment.
[9] See, for example, Bellemare and Kröger (2007) and Falk and Zehnder (2013).

half of them are male (54%), and a majority identify as Christians (67%), specifically Catholics or Protestants. In terms of socioeconomic background, most students come from families falling into one of the following categories: (i) two-parent households where at least one parent has a high level of education (20%); (ii) two-parent households with both parents having a low level of education (27%); and (iii) single-parent households with a parent who has a low level of education (27%).

# 3 Empirical Approach

## 3.1 Identification

We are interested in estimating how the ethnic composition of classroom peers affects native students' in-group out-group bias. As pointed out by Manski (1993), two challenges in identifying peer effects are correlated unobservables and endogenous group formation. Translated to our setting, the first challenge is that the ethnic makeup of a classroom is likely correlated with both observable and unobservable characteristics, such as average family income and attitudes towards immigrants. The second challenge is that students (and their families) self-select into the neighborhoods and schools they attend in ways which are likely to create a bias.

To deal with these challenges, we take advantage of the quasi-random assignment of students to classrooms within schools. The idea is that while the school a student attends is unlikely to be random, which classroom they are assigned to within a school is as a good as random. We model outcomes (e.g., in-group bias) for native individual $i$ in classroom $k$ in school $s$ as:

$$Y_{i,k,s} = \alpha + f(diversity_k) + \delta \mathrm{X}_i + \gamma \mathrm{Z}_k + \theta_s + \epsilon_{i,k,s} \tag{1}$$

where $f(\cdot)$ is a function of the mix of ethnic peers ($diversity_k$) a native is exposed to in their classroom. We will model classroom diversity either as a function of the fraction of immigrant peers or as a function of the fraction of culturally distant and the fraction of culturally close immigrant peers. The vector $X_i$ contains individual and family background characteristics, $Z_k$ is the number of students in the class, and $\epsilon_{i,k,s}$ is the error term. Crucially, the estimating equation includes school fixed effects, $\theta_s$, to account for the fact that individuals are not randomly assigned to schools.[10]

Figure 3 illustrates the identifying variation we use when estimating our model. The left panel displays a histogram for the fraction of immigrant peers in classes, where we have first regressed out the school fixed effects. The scatterplot in the right panel likewise shows the mix of Muslim versus non-Muslim immigrant peers in a classroom after netting out school fixed effects. Both graphs reveal substantial residual variation in our diversity measures.

Our identification strategy is related to work which exploits either within school variation across

[10]Note that all but one of our outcome variables are calculated as within-person differences. For example, the in-group bias is the difference in how much a native invests in a native versus how much they invest in an immigrant. This specification nets out the constant effect within an individual, which provides an increase in precision.

classes or natural variation in cohort composition across time within a given school (Antecol et al., 2015; Hoxby, 2002; Hanushek et al., 2003; Carrell and Hoekstra, 2199; Carrell et al., 2018; Balestra et al., 2022).[11]

### 3.2 Validity

Our identification strategy relies on quasi-random assignment of students to classes within a school. In Germany, all students have the right to receive a non-discriminatory education, and schools often have explicit rules to ensure this applies in the assignment of students to classrooms.

The right to a non-discriminatory education evolved over time. In 1971, the Standing Conference of the Ministers of Education and Cultural Affairs of the German Federal States issued a guideline that immigrants should be treated equally in all matters related to schooling (Puskeppeleit and Krüger-Potratz, 1999). Despite this, up to the 1990's, classrooms were often partially segregated along native versus immigrant lines. A major citizenship reform introducing birthright citizenship in 2000 coincided with a renewed push to integrate classrooms (Nieden and Karakayali, 2016).

In a high profile case in Berlin in 2012, German parents lobbied for and were successful at creating classrooms which were highly segregated. Immigrant parents filed a complaint with the Berlin Senate arguing this was a discriminatory practice and the Senate ruled in their favor, requiring students to be reassigned in a non-segregated fashion (Nieden and Karakayali, 2016). Today, all German federal states have school laws that specify a right to non-discriminatory education (Federal Anti-Discrimination Agency, 2022).

Although these laws do not explicitly mandate random assignment, in practice many schools state that they do not discriminate based on migration background when assigning students to classrooms.[12] In our own conversations with school officials and principals, we were also told that migration background is not used as a criterion for making classroom assignments. Consistent with this, a recent government report highlights that in Germany "the assignment of students to classes is difficult to influence" (Federal Government of North Rhine-Westphalia, 2022, p. 92).[13]

Several empirical tests provide strong support that students are quasi-randomly assigned to classrooms. Our first test appears in Table 2. We regress the fraction of immigrant peers in a native's class on background characteristics of the native student and the number of students in the class. In column 1, we do not include school fixed effects. Several variables have sizable and statistically significant effects. In particular, native students with a higher fraction of immigrant peers are less likely to be Protestant or non-religious, are more likely to come from families with lower

[11]Another approach uses the random assignment of peers to social groups. Prominent examples are the random assignment of students to classrooms, such as in project STAR (Chetty et al., 2011), roommates in university dorms (Sacerdote, 2001), freshmen to university sections (Feld and Zölitz, 2017), and cadets to squads in the military (Lyle, 2007; Dahl et al., 2021).

[12]See, for example, `https://www.goetheschule-asperg.de/index.php/eltern/faq` (accessed November 14, 2023).

[13]There is limited scope to influence assignment; for example, while some schools allow parents to name one friend they would like to have their child be in the same class with, others state this type of preference will not be considered.

socioeconomic backgrounds, and are older. An F-test reveals that the variables are also jointly significant (p-value<.001). Column 2 of Table 2 reports what happens when school fixed effects are added to the regression. All of the coefficient estimates are now close to zero, with none of the 15 variables being individually statistically significant. The joint F-test is not statistically significant either (p-value=.329). The tests in column 2 are similar to the balancing tests performed for actual experiments, to check whether random assignment to treatment has been implemented correctly.

In columns 3 and 4 of Table 2, we perform a similar test, but this time using the fraction of Muslim immigrant peers as the dependent variable. As before, when school fixed effects are not included, the covariates are both individually and jointly statistically significant. But when school fixed effects are added to the regression, only one of the coefficients is significant at the 10% level (roughly what would be expected by chance) and the covariates are not jointly significant (p-value=.409).

Related to these balancing tests, for our main regression model in Section 4, we explore what happens when we add additional covariates beyond the school fixed effects. As expected with conditional random assignment, the estimates are virtually identical.

Finally, we conduct a simulation test. Following Carrell and West (2010), we randomly assign students to classrooms within schools, using the actual share of immigrants at the school level and the actual class sizes within a school. We repeat this counterfactual exercise 1,000 times. To test random assignment, for every set of re-sampled classes, we calculate the empirical p-value as the proportion of simulations where exposure to immigrant students is smaller than that observed in the original class. If class composition is random, the distribution of p-values within a school should be approximately uniform, which is testable using a one-sample Kolmogorov-Smirnov test. We reject uniformity for just one out of 57 schools in our sample at the 5% confidence level, and obtain the same result (rejecting uniformity just once) when repeating the simulation exercise for the share of Muslim immigrants.

In conclusion, the battery of tests we conduct all provide strong support for the quasi-random assignment of immigrant and native students to classrooms.

### 3.3 Modeling Classroom Diversity

To model classsroom diversity, $f(diversity_k)$, we build on the measures of Montalvo and Reynal-Querol (2005) which distinguish between ethnic polarization and ethnic fractionalization:

$$polar_k = 4\sum_{j=1}^{N} \pi_{j,k}^2(1-\pi_{j,k}) \quad \text{and} \quad frac_k = 2\sum_{j=1}^{N} \pi_{j,k}(1-\pi_{j,k}), \tag{2}$$

where $\pi_{j,k}$ represents the proportion of individuals from ethnic group $j$ in classroom $k$. The polarization index reaches its maximum value when a classroom is characterized by a bipolar composition, meaning it consists of only two large and equally sized groups. Conversely, the fractionalization index attains its maximum value when the proportions of each of the $N$ groups within the classroom are equal, each accounting for $1/N$ of the total population.

In the first step of our analysis, we view classrooms as being composed of two groups: natives (with share $1 - \pi_k$) and immigrants (with share $\pi_k$). In the case of only two groups, where $polar_k = 4\pi_k(1-\pi_k)$ and $frac_k = 2\pi_k(1-\pi_k)$, the indices are equivalent up to a scaling factor, both reaching their maximum when natives and immigrants constitute equally large groups. Therefore, we start by modeling $f(diversity_k)$ in equation 1 as a quadratic polynomial, $\beta_1\pi_k + \beta_2\pi_k^2$, where $\pi_k$ denotes the share of immigrant classmates. Our first regression specification is:

$$Y_{i,k,s} = \alpha + \beta_1\pi_k + \beta_2\pi_k^2 + \delta \mathrm{X}_i + \gamma \mathrm{Z}_k + \theta_s + \epsilon_{i,k,s} \tag{3}$$

In the second step of our analysis, we consider two immigrant subgroups within classrooms: immigrants culturally close to their native peers and culturally distant immigrant peers. As laid out in Section 2.2, our main proxy for cultural distance is based on religion, specifically the distinction between non-Muslim and Muslim immigrants. Thus, a classroom is now defined as being composed of non-Muslim immigrants (share $\pi_{C,k}$, with $C$ for culturally "close"), Muslim immigrants (share $\pi_{D,k}$, with $D$ for culturally "distant"), and native Germans (share $\pi_{N,k}$, with $N$ for "natives"). In this case, using $\pi_{N,k} = 1 - \pi_{C,k} - \pi_{D,k}$, we can express $polar_k$ and $frac_k$ as functions of the share of non-Muslim immigrant peers ($\pi_{C,k}$) and the share of Muslim immigrant peers ($\pi_{D,k}$):

$$f(diversity_k) = \underbrace{\overbrace{4(\pi_{C,k} + \pi_{D,k} - \pi_{C,k}^2 - \pi_{D,k}^2 - \pi_{C,k}\pi_{D,k})}^{=frac_k} - 12(\pi_{C,k}\pi_{D,k} - \pi_{C,k}^2\pi_{D,k} - \pi_{C,k}\pi_{D,k}^2)}_{=polar_k} \tag{4}$$

The first term in brackets is the expression for the fractionalization index. The sum of the two terms on the right hand side is the reformulated polarization index. Thus, our second regression specification, which flexibly nests both polarization and fractionalization, is:

$$\begin{aligned} Y_{i,k,s} = &\alpha + \beta_1\pi_{C,k} + \beta_2\pi_{D,k} + \beta_3\pi_{C,k}^2 + \beta_4\pi_{D,k}^2 + \beta_5\pi_{C,k}\pi_{D,k} + \\ &\beta_6\pi_{C,k}^2\pi_{D,k} + \beta_7\pi_{D,k}^2\pi_{C,k} + \delta \mathrm{X}_i + \gamma \mathrm{Z}_k + \theta_s + \epsilon_{i,k,s} \end{aligned} \tag{5}$$

Note that the specification of classroom diversity in equation 5 is close to a third order expansion of the terms $\pi_{C,k}$ and $\pi_{D,k}$. The only difference is that a third-order expansion would have also included $\pi_{C,k}^3$ and $\pi_{D,k}^3$.

In an extension in Section 4.3, we will also consider more than two immigrant groups based on the parents' country or region of origin. For this extension, we model diversity using the expressions for polarization and fractionalization in equation 2.

# 4 Results

## 4.1 Share of Immigrant Peers in the Classroom: An Inverted-U Relationship

We start by examining how natives' in-group bias varies with the proportion of immigrant peers in a classroom in Table 3. The dependent variable is the in-group out-group investment gap. For comparison purposes, we start with a linear specification in panel A. In column 1, we include only basic controls: student's gender, age, the class size, and school fixed effects. There is no statistically significant linear relationship between the in-group bias of native students and the proportion of immigrant peers. This null effect continues to hold when we additionally include controls for students' religious background (column 2) and family background (column 3). Based on panel A, one might be tempted to conclude that in-group bias is unaffected by classroom diversity, but this would be incorrect, as panel B demonstrates.

In panel B, we model diversity as a quadratic function of the proportion of immigrant peers, as specified in equation 3. No matter which set of controls are included, the coefficients on the polynomial terms are individually and jointly significant (p-value for joint F-test = .0001 in panel B, column 3). Moreover, we cannot reject that the peak in in-group bias occurs when polarization is highest (50% natives and 50% immigrants).[14]

Figure 4 illustrates this inverted-U shaped relationship. The x-axis is the fraction of immigrant peers in a classroom and the y-axis is the investment gap in standard deviation units. The in-group out-group investment gap initially widens as the proportion of immigrant classmates increases, but then it reverses course and starts to narrow again as the share of immigrant peers continues to rise. At the turning point of a 45% immigrant share, a one-standard deviation change in the immigrant share in either direction (0.21) results in a statistically significant 10% of a standard deviation reduction in in-group bias (p-value=0.018). A similar pattern emerges when estimating diversity even more flexibly by using a third-order polynomial in the fraction of immigrant peers (see Appendix Figure A1).

## 4.2 Diversity Among Immigrant Peers and the Role of Cultural Distance

Our first analysis modeled classroom diversity solely as a function of the proportion of immigrants in a classroom, treating immigrants as a homogeneous group. A natural question is whether diversity within the immigrant group matters. Once there are more than two groups (natives versus immigrants), the distinction between polarization and fractionalization also becomes relevant. This is because the expressions for polarization and fractionalization in equation 2 are equivalent up to a scaling factor when there are just two groups, but differ in the case of three groups or more (equation 4).

[14]The test is whether the two quadratic coefficients are equal but opposite in sign; this test has a p-value of .359 in panel B, column 3.

With this in mind, we now shift our focus to exploring how diversity among immigrants matters, distinguishing between immigrants with different cultural backgrounds. Using the three group case, we ask the following questions. What role does cultural diversity within the immigrant group play in shaping the inverted-U relationship between natives' in-group bias and the immigrant share? Is this relationship due to a high level of polarization, where students are exposed to a two-point symmetric distribution of classmates, with natives being equal in number to one large and culturally distinct group of immigrants? Or is it explained by fractionalization, where natives, culturally close immigrants, and culturally distant immigrants are each minorities?

In Germany, perhaps the most salient measure for whether an immigrant is culturally close or distant to a native German is captured by religious affiliation—specifically, whether an immigrant is Muslim. Our second analysis uses the share of Muslim and non-Muslim peers to flexibly model the effects of classroom diversity using equation 5. The results can be found in Table 4. The estimates describe a three-dimensional surface where the x-axis is the share of culturally distant peers, the y-axis is to the share of culturally close peers, and the z-axis is the predicted in-group bias among natives (as measured by the in-group out-group investment gap).

Instead of plotting three-dimensional figures, we show heatmaps, as they make it easier to visualize our results. A two-dimensional heatmap plots the predicted values of the in-group out-group investment gap using colors, with darker shading representing higher values and lighter shading representing lower values. The heatmap depicted in Figure 5 corresponds to the estimates appearing in Table 4.[15]

The peak in native's in-group bias is observed in classrooms marked by what we define as *cultural polarization (CP)*. This occurs when native German peers constitute a slight majority of 60%, while Muslim immigrant peers make up a large minority of 40%. The table displayed to the right of the heatmap in Figure 5 calculates the difference between predicted in-group bias at its peak ($CP$) and four other classroom scenarios. Each of these scenarios are also labeled on the heatmap.

In the first scenario, referred to as *non-cultural polarization (NCP)*, native Germans continue to make up a slight majority of 60%, but now the non-Muslim immigrants are the large minority group of 40%. In comparison to the peak in in-group bias found in culturally polarized classrooms, native students in these classrooms exhibit 33% of a standard deviation lower in-group bias.[16] This can be seen visually in the heatmap by the darker color shading at $CP$ versus $NCP$. In the second scenario, which we term *mixed polarization (MP)*, native peers once again constitute a slight majority of 60%, while both Muslim and non-Muslim immigrant peers now form two medium-sized minority groups, each comprising 20% of the students in a class. In comparison to the peak in culturally polarized classrooms, the in-group bias of native students is 32% of a standard deviation lower in

[15]The range for the heatmap is limited to the areas where we have nontrivial amounts of data: non-Muslim immigrant shares between 0% to 50%, Muslim immigrant shares between 0% to 80%, and the combined total of these shares not exceeding 80%. All control variables are evaluated at their means.

[16]For context, one standard deviation equals 76 euro cents and the average in-group out-group bias is 9 euro cents, so this represents a 279% increase relative to the mean.

this setting. Revisiting the inverted-U relationship between the in-group bias of native students and the fraction of immigrant peers, these findings imply the peak is primarily driven by culturally polarized classrooms. In contrast, non-cultural and mixed polarization seem to alleviate in-group biases among native students.

The finding that culturally polarized classrooms drive differences in in-group bias gains further support when we consider two additional classroom scenarios. The first, labeled *overall fractionalization (OF)*, evenly divides the classroom among native peers, Muslim immigrant peers, and non-Muslim immigrant peers, with each group constituting one-third of classmates. The second, denoted as *native supermajority (NSM)*, features native peers as a large majority of 80%, with both Muslim and non-Muslim immigrant peers forming small minority groups of 10% each. In both of these scenarios, the in-group bias of native students is approximately 40% of standard deviation lower compared to culturally polarized classrooms.

Our next result highlights that natives in culturally polarized classrooms experience payoff losses due to their in-group bias. For each native, we calculate the expected difference in their payoffs as the sender (first mover) if they are randomly matched with an immigrant versus a native interaction partner (second mover).[17] Using the expected payoff difference as the outcome variable, we estimate equation 5.

The resulting heatmap is displayed in Figure 6. The payoff loss reaches a peak when there are 55% natives and 45% Muslim immigrants in a classroom. In comparison, the peak in in-group bias shown in Figure 5 occurs at a similar point (60% natives and 40% Muslim peers). More generally, comparing the two heatmaps, they have almost identical patterns. To test this more formally, we compare the predicted payoff losses in culturally polarized classrooms versus the four other classroom types. Importantly, classroom types are defined based on the definitions used in Figure 5; we do not redefine what a culturally polarized classroom is to reflect the peak in payoff losses in Figure 6.

We find larger, and statistically significant, payoff losses in culturally polarized classrooms relative to other classroom types. Payoff losses are 34% of a standard deviation higher compared to non-culturally polarized classrooms. Likewise, payoff losses are 30 to 44% of a standard deviation higher relative to classrooms with mixed polarization, overall fractionalization, or native supermajorities. Thus, the pronounced in-group biases in culturally polarized classrooms reflects behavior that results in financial losses for natives.

[17] As a reminder, all individuals fill out decisions sheets as the first mover playing with a native and with an immigrant as well decision sheets as the second mover playing with a native and with an immigrant. Participants are told they will be randomly chosen to play either as the first mover or the second mover and that they will be paired randomly with a native or an immigrant.

### 4.3 Robustness and Extensions

As a first robustness test, we augment regression equation 5. As a reminder, equation 5 is a flexible nesting of both the polarization and fractionalization indices. It is close to a third order expansion of the non-Muslim and Muslim immigrant fractions in a classroom, but does not include the two cubics $\pi^3_{C,k}$ and $\pi^3_{D,k}$. When we add these terms into the regresssion, the resulting heatmap is similar (see Appendix Figure A2).

As a second robustness check, we expand the number of immigrant groups based on parents' country of origin.[18] Countries of origin are mapped into 11 regions: Turkey, Balkan States, Eastern Europe, Post Soviet Bloc, Southern Europe, Central and Northern Europe, Middle East, Asia, Africa, other countries, and unidentified (see Appendix Table A1).

Since we now have 11 immigrant groups instead of two, we estimate a more parsimonious model for diversity. Specifically, we use the summary polarization and fractionalization measures in equation 2 as two right-hand side variables. Appendix Figure A3 illustrates that there is independent variation in these two variables in the 11 group case after netting out classroom fixed effects. The figure plots a classroom's residualized polarization on the x-axis and residualized fractionalization on the y-axis. Holding constant polarization, there can be sizable differences in fractionalization and *vice versa*.

Table 5 contains the in-group out-group investment gap estimates for the 11 immigrant group case. The first column only includes the polarization measure, and finds a large coefficient. The second column only includes the fractionalization measure, and also finds a large estimated coefficient. But when both measures are included in column 3, only the polarization variable matters. The polarization coefficient is large and statistically significant; the estimated effect of 0.54 implies that a one standard deviation increase in polarization (0.147) leads to 8% of a standard deviation increase in in-group bias. In sharp contrast, the fractionalization coefficient is close to zero and insignificant. Of course, this more parsimonious model does not distinguish between different types of polarization.

## 5 Mechanisms

In our experiment, participants are paired with native and immigrant interaction partners *drawn from the general student population* (i.e., not from students in their own classroom or school). In these interactions, natives in culturally polarized classrooms exhibit large in-group biases and lose money because of it. We now explore why this might be the case.

Two important rationales for transferring money in the investment game are trust and other-regarding preferences. Trust refers to the sender's beliefs about the recipient's trustworthiness and their expectation of generous reciprocation. Other-regarding preferences encompass concerns for the well-

[18] If both parents are immigrants, we use the mother's origin country; if the mother is an immigrant and the father is not, we use the mother's country; if the father is an immigrant and the mother is not, we use the father's country.

being of others, driven by considerations like altruism or fairness. Based on this, our main hypothesis has two parts. First, natives in culturally polarized classrooms could develop more mistrust toward immigrants because of *limited or adversarial interactions* with immigrants in their classroom. Second, if natives *generalize* these experiences, they may form negative stereotypes, leading to reduced trust and lower other-regarding preferences toward immigrants in general. Consequently, they would rationally favor natives over immigrants in general, based on these biased beliefs or preferences. To address both parts of this hypothesis, we explore (i) the formation of friendships and shared identities between natives and immigrants within classrooms as well as peers' behavior, and (ii) whether natives indeed generalize their classroom experiences to the broader immigrant population.

## 5.1 Within the Classroom: Intergroup Friendships, Shared Identities, and Peers' Behavior

**Intergroup Friendships.** We begin by empirically examining how different forms of classroom diversity influence the formation of friendships between natives and immigrants. We construct an index of intergroup friendships using our survey, which asks participants "What does your circle of friends look like: How many of your friends are not from Germany themselves, or have parents who are not from Germany?" Respondents could answer on a scale of 1 to 6, where 1=none, 2=less than one quarter, 3=roughly one quarter, 4=roughly one half, 5=most, and 6=all. We assign shares of 0%, 12.5%, 25%, 50%, 75%, and 100% to the answers of 1-6, respectively.

One limitation of our survey question is that it captures the proportion of immigrant friends in general, not just within the classroom. However, by comparing the survey responses of native and immigrant students within the same class, we can assess the extent to which their friendship networks potentially overlap. Specifically, our intergroup friendship index draws on the concept of homophily from network science and quantifies how much the friendships of native and immigrant students in a class deviate from what would be expected if friendships were formed randomly. Asymptotically, random formation would lead both groups to have the same proportion of immigrant friends. However, in finite-sized classrooms, natives are expected to have a slightly higher proportion of immigrant friends than immigrants, since an immigrant cannot befriend themself. The expected difference in proportions is $(N_{k,s} - 1)^{-1}$, where $N_{k,s}$ represents the size of class $k$ in school $s$.[19]

Our index measures whether a native student $i$ in classroom $k$ in school $s$ has a lower or higher proportion of immigrant friends compared to immigrant classmates, relative to the random formation benchmark. Formally, it is defined as:

$$Y_{i,k,s} = \frac{f_{i,k,s} - \bar{f}_{immig,k,s}}{(N_{k,s} - 1)^{-1}}. \tag{6}$$

[19]To see this, let $\pi_k$ denote the proportion of immigrants in a given class $k$. Under random friendship formation, the expected proportion of immigrant friends for native students would be $N_{k,s}\pi_k/(N_{k,s} - 1)$, while the corresponding proportion for immigrant students would be $(N_{k,s}\pi_k - 1)/(N_{k,s} - 1)$. The difference between these two proportions is $(N_{k,s} - 1)^{-1}$.

The numerator calculates the difference between the share of friends with an immigrant background reported by each native respondent ($f_{i,k,s}$) and the average share of immigrant friends reported by all immigrants in the same class ($\bar{f}_{immig,k,s}$). A value of the index below unity indicates fewer cross-group friendships than would be expected under random formation (i.e., homophily); this includes cases where native students report a lower proportion of immigrant friends than immigrants classmates do. Conversely, a value above unity implies stronger than expected cross-group friendships (i.e., heterophily).

We estimate equation 5 with the intergroup friendship index, normalized to be mean zero and standard deviation one, as the dependent variable. Figure 7 with its heat map and corresponding table presents the key findings. In line with our hypothesis, classrooms with approximately equal proportions of natives and Muslim immigrants tend to hinder the formation of intergroup friendships. The intergroup friendship index reaches its lowest value when natives represent 44% and Muslim immigrants 56%. This point of minimal intergroup friendships is slightly shifted to the right compared to the peak of in-group bias observed in Figure 5, marked by $CP$ in the current heatmap. Despite this shift, the overall patterns in the two heatmaps remain strikingly similar.

This becomes even more evident when comparing intergroup friendship formation in culturally polarized classrooms to other classroom types. As before, we continue to use the same definitions for classroom types from Figure 5 for in-group bias; we do not redefine culturally polarized classrooms based on the intergroup friendship minimum. The results show substantial and statistically significant declines in the intergroup friendship index for culturally polarized classrooms compared to each of the four other classroom configurations. Specifically, the friendship index is 104% of a standard deviation lower than in non-culturally polarized classrooms, 54% lower than in mixed polarized classrooms, 55% lower than in overall fractionalized classrooms, and 85% lower than in native supermajority classrooms.

**Shared Identities.** Another testable implication of our hypothesis is that natives and immigrants in culturally polarized classrooms are unlikely to develop shared identities, as such an environment hinders social interactions between them.

To empirically investigate this, we create an index that measures the difference in how strongly natives and their immigrant classmates self-identify as German. In our survey, we asked participants, "Generally speaking, how much do you identify as German?" Responses were recorded on a scale from 1 to 5, where 1=not at all, 2=barely, 3=in some ways, 4=mostly, and 5=fully. Our identity gap index quantifies the difference between each native student's level of self-identification and the average level of self-identification among their immigrant classmates. We then estimate equation 5 using this identity gap as the dependent variable.

Figure 8 presents the heatmap for the identity gap index, showing that the gap between natives and immigrants is largest when a classroom is composed of 55% natives and 45% Muslim immigrants. This closely corresponds with the peak for in-group bias seen in Figure 5. The identity gap in culturally polarized classrooms is 94% of a standard deviation higher compared to non-culturally polarized classrooms. Similarly, there are substantial and statistically significant differences when compared to mixed polarized, overall fractionalized, and native supermajority classrooms.

**Peers' Behavior.** As a final test for how different forms of diversity play out inside classrooms, we use information from the other side of the experiment, where participants play as the receiver and are asked how much money they would return based on whether the sender was a native or immigrant.

We calculate how natives would have been treated by natives versus immigrants had they been paired with someone from their own classroom. This is a counterfactual exercise, as students were specifically told they would not be paired with a classmate, but rather someone from a different school. But it still serves as a proxy measure for the differential treatment a native is likely to experience in their actual classroom, which they may mistakenly extrapolate to the broader population.

We measure differential treatment by calculating how much, on average, native and immigrant receivers transfer back to native senders. As a reminder, receivers specify how much they would transfer back for each of 11 possible investments (50 cent increments from 0 to 5 euros). Therefore, we average the back transfers for native and immigrant receivers over the 11 possible investments from a native sender. This approach holds constant the amount received from a native sender, enabling a direct comparison.[20]

We estimate equation 5 using this measure for the native-immigrant gap in reciprocity as the dependent variable, and illustrate the main results in Figure 9 and the accompanying table. The largest reciprocity gap occurs in culturally polarized classroom. Relative to Figure 5, the current graph has the same general pattern, but with more pronounced differences. We find statistically significant increases in the reciprocity gap in culturally polarized classrooms which are between 114% to 121% of a standard deviation higher compared to non-culturally polarized, fractionally polarized, overall fractionalized, and native supermajority classrooms.

### 5.2 Outside the Classroom: Trust and Other-Regarding Preferences towards the Broader Immigrant Population

**Trust.** We have just presented evidence indicating that natives in culturally polarized classrooms form fewer intergroup friendships, have less of a shared identity, and are treated less favorably by immigrants than by natives in their class. If these classroom experiences are generalized to the broader immigrant population, natives may develop negative stereotypes, leading to reduced trust

[20] While we also asked senders how much they would expect to get back if they played against an immigrant versus a native, this does not allow for a clean comparison. The reason is that these expectations were asked unconditionally. Hence, a native sender could partly expect to get back more from natives versus immigrants because they chose to invest more in natives versus immigrants.

in *all* immigrants compared to natives. Consequently, they might rationally choose to make larger transfers to natives over immigrants, driven by these biased beliefs. This could help explain the significant in-group biases observed in culturally polarized classrooms.

To test this, we use questions in our survey about generalized trust. Specifically, we asked participants how much do you trust people with German nationality and how much do you trust people with a foreign nationality. Students could answer on a scale from 0 to 10, with a 0 indicating "very low degree of trust" and a 10 "very high degree of trust". These same two questions are asked on the Eurobarometer.

Using the within subject difference in trust (trust in natives - trust in immigrants) as the outcome variable, we estimate equation 5. Figure 10 displays the associated heatmap. The largest in-group out-group trust gap occurs with 57% natives and 43% Muslim immigrant peers. As shown in the graph, this peak closely aligns with the peak in cultural polarization defined in Table 5 and marked in the current figure with the label $CP$. More generally, the two heat plots have very similar patterns.

We find statistically significant increases in the trust gap in culturally polarized classrooms relative to each of the other four classroom configurations. The differences are large. The trust gap is 50% of a standard deviation higher in culturally polarized classrooms relative to non-culturally polarized classrooms, 33% higher relative to mixed polarized classrooms, 38% higher in comparison to overall fractionalized classrooms, and 43% higher relative to native supermajority classrooms.

**Other-Regarding Preferences.** A generalization of classroom experiences could also influence social preferences, leading natives in culturally polarized classrooms to favor making larger transfers to other natives over immigrants. Other-regarding preferences which have an in-group bias is conceptually distinct from the belief that natives are more trustworthy and therefore more likely to reciprocate generously. We investigate this possibility through two different tests.

We begin by again using information from the other side of the experiment, but this time we focus solely on the decisions natives make when they play as the receiver. Native receivers are asked how much money they would return based on both the amount of the transfer and whether the sender was a native or immigrant. Note that when natives play as the receiver (second mover), they have no financial incentive to return any money to the sender (first mover). This is because they are playing a one-shot game, where the senders and receivers are completely anonymous and come from different schools.

For these reasons, the choices made by a receiver in our game are akin to those made by dictators in a dictatorship game. In a one-shot dictator game against an anonymous opponent, a self-interested dictator's optimal strategy is to keep all of the money and return nothing. If they do return money, this can be attributed to other-regarding preferences. The incentives faced by a self-interested receiver in our investment game are similar. And since the receiver conditions their response on whether they are playing against a native or an immigrant, we can measure a native receiver's difference in other-regarding preferences.

Specifically, we measure differences in other-regarding preferences by calculating how much, on average, native receivers transfer back to native versus immigrant senders. As a reminder, receivers specify how much they would transfer for each of 11 possible investments (50 cent increments from 0 to 5 euros). Therefore, we average native receivers back transfers over the 11 possible investments, enabling a direct comparison. We then estimate equation 5 using this as the dependent variable.

Figure 11 displays the heatmap associated with this measure of social preferences. Natives show the strongest preference for making larger transfers back to fellow natives rather than to immigrants when the classroom consists of 68% natives and 32% Muslim immigrants. This maximum is shifted somewhat to the left compared to the peak in cultural polarization for in-group bias observed in Figure 5 and marked by $CP$ in the figure. Nevertheless there is a similar pattern in the two heatmaps. The differences between culturally polarized classrooms and the four other classroom types are all positive, statistically significant, and large (between 34 and 49% of a standard deviation).

As a second test for social preferences, we construct a measure of anti-immigrant sentiment among natives. Survey participants were asked whether it is fair that workers of Turkish descent are allowed to work in Germany and similar questions for immigrants of Polish and French origins. They could answer on a scale of 1 to 4, where 1=strongly agree, 2=agree somewhat, 3=disagree somewhat, and 4=strongly disagree. Using a principal components analysis, we construct an index of anti-immigrant sentiment based on these three questions, normalizing the index to have mean 0 and standard deviation 1. A larger value indicates a native has more anti-immigrant sentiment. Note that this outcome is somewhat different from all of the other outcomes used in the paper, as it is measured in levels rather than as a difference.

The results for anti-immigrant sentiment is shown in Figure 12. The peak occurs close to the peak in cultural polarization for in-group bias, as marked with the label $CP$ in the graph. And as we have seen in the many heatmaps preceding this one, the pattern is similar to that observed in Figure 5. The difference in anti-immigrant sentiment for a culturally polarized classroom versus non-culturally polarized classroom is 40% of a standard deviation. There are likewise large and statistically significant differences relative to mixed polarized, overall fractionalized, and native supermajority classrooms. This provides evidence that negative attitudes towards the broader immigrant population contributes to the in-group bias observed in Figure 5.

### 5.3 Relationship to Discrimination

As mechanisms for the observed in-group bias in culturally polarized classrooms, we found evidence that natives had limited and negative interactions with immigrants in their classroom and that they generalized these experiences to the broader population of immigrants. Here we briefly discuss how these findings can be understood through the perspective of different forms of discrimination.

By design, our experiment rules out the possibility of (accurate) statistical discrimination playing a role. In our setting, statistical discrimination would be a rational perception among natives that the immigrants they play with will reciprocate less generously than natives. We set up our experiment so that any statistical discrimination would be constant across classroom types. Specifically, students were told they would engage in the investment game with an individual (native or immigrant) chosen randomly from a different school in their federal state of residence, rather than someone from their own class. Since all natives play with the same set of immigrants on average, there is no role for differential statistical discrimination across classrooms with different types of diversity. Thus, it cannot explain the peak in in-group bias observed in culturally polarized classrooms versus other classroom scenarios.

Since statistical discrimination is ruled out by design, the remaining explanations are negative stereotypes and taste discrimination, both of which are costly. Indeed, these two explanations are consistent with the payoff losses observed in culturally polarized classrooms in Figure 6.

Negative stereotypes represent *inaccurate* statistical discrimination (Bohren et al., forthcoming), where natives in culturally polarized classrooms develop incorrect beliefs about how immigrants in the broader population will treat them based on their in-class interactions and experiences. The low levels of intergroup friendship and shared identity in culturally polarized classrooms observed in Figures 7 and 8 could contribute to negative stereotypes. Likewise, natives could mistakenly extrapolate the larger native-immigrant gap in reciprocity in culturally polarized classrooms found in Figure 9 to the broader population. These negative stereotypes would naturally lead to lower levels of both trust in immigrants (Figure 10) and other-regarding preferences for immigrants in the broader population (Figures 11 and 12).

Taste discrimination occurs when natives, due to their social preferences, favor natives over immigrants. Our two results for other-regarding preferences are also consistent with this explanation. Native receivers prefer to transfer back more money to natives versus immigrants in culturally polarized classrooms (Figure 11), in a game where there is not a financial incentive to return money to a sender. Likewise, anti-immigrant sentiment is largest in culturally polarized classrooms, as seen in Figure 12. Of course, taste discrimination and negative stereotypes could interact with each other. In particular, an increase in negative stereotypes could affect an individual's preferences, amplifying their taste discrimination. And taste discrimination could contribute to negative stereotypes.

# 6 A Framework

In this section, we posit a simple model that rationalizes our empirical findings. The model features individuals who actively choose who to identify and socially engage with. A lack of identification with out-group members would naturally explain the in-group biases we observe, as well as our evidence on mechanisms. Individuals are of different types—e.g., natives and multiple immigrant types—and these types differ in terms of their endowed cultural identity. They must decide whether to join an *inclusive* group with a shared identity or preserve their innate cultural identity by joining an *exclusive* group comprised solely of others of their own type. This decision involves striking a balance between the wish to be part of a larger group and the willingness to forfeit one's own cultural identity in favor of that of the inclusive group, where the latter identity shifts endogenously with the distribution of types in the population. Individual heterogeneity is introduced through additive extreme value-distributed preferences for inclusive versus exclusive group membership. Given that each individual faces a binary decision and cares about how many others make the same choice—that is, join the same group—our model can be seen as a version of the social interactions model by Brock and Durlauf (2001), but with an in-built Hotelling structure where multiple interacting types are endowed with different cultural identities. In our framework, natives' in-group biased behavior is naturally reflected by the proportion of natives who actively choose not to engage with immigrants in the inclusive group but instead choose to engage exclusively with other natives.

**Basic Structure.** Consider an economy with a continuum of individuals who are of $J \geq 2$ types. Let the proportions of types in the population be denoted $\boldsymbol{\pi} \equiv (\pi_1, ..., \pi_J)$. There is a space of cultural identities which we take to be $\Theta \equiv [-1, 1]^{J-1}$. Each individual $i$ of type $j$ is endowed with some exogenously given identity $\boldsymbol{\theta}_j \in \Theta$ and the inclusive group has an endogenous identity $\boldsymbol{\theta} \in \Theta$, making the type-specific (Euclidean) distance $d_j \equiv \|\boldsymbol{\theta} - \boldsymbol{\theta}_j\|$. We assume a type-specific cost of joining the inclusive group that depends on this cultural distance, $h_j = h_j(d_j)$ where, for each type $j$, $h_j(\cdot)$ is twice continuously differentiable and satisfies $h_j(0) = 0$, $h'_j(\cdot) > 0$ and $h''_j(\cdot) > 0$. Finally, let $\beta > 0$ parameterize the strength of preference for group size. Individual $i$ of type $j$ chooses between two options: either stay with the type-$j$ exclusive group (option 0) or join the inclusive group (option 1). Let $\mu_j \in [0, 1]$ denote the proportion of type $j$ individuals who choose to join the inclusive group. The associated utilities for individual $i$ are then

$$u^0_{ij} = \beta\pi_j (1 - \mu_j) + \varepsilon^0_{ij}, \qquad u^1_{ij} = \beta \sum_{j'=1}^{J} \pi_{j'}\mu_{j'} - h_j + \varepsilon^1_{ij}, \qquad j = 1, ..., J,$$

where we used that the size of the type-$j$ exclusive group is $\pi_j (1 - \mu_j)$ and size of the inclusive group is $\sum_{j'=1}^{J} \pi_{j'}\mu_{j'}$, and where $\varepsilon^0_{ij}$ and $\varepsilon^1_{ij}$ are the individual's i.i.d. extreme value distributed choice-specific random preferences. For any given type distribution $\boldsymbol{\pi}$ and inclusive group identity $\boldsymbol{\theta}$ (and hence fixed vector of joining costs $\boldsymbol{h} \equiv (h_1, ..., h_J)$), a "joining equilibrium" is a vector of type-specific inclusive-group joining rates $\boldsymbol{\mu} \equiv (\mu_1, ..., \mu_J)$ that satisfies

$$\log\left(\frac{\mu_j}{1 - \mu_j}\right) = \beta\left(\sum_{j'=1}^{J} \pi_{j'}\mu_{j'} - \pi_j (1 - \mu_j)\right) - h_j, \qquad j = 1, ..., J,$$

for all $J$ types simultaneously. In Appendix B, we prove that such a joining equilibrium exists and is guaranteed to be unique for $\beta \in (0, 2)$ for any finite number of types $J$. The upper limit, $\beta \leq 2$, ensures the absence of multiple coordination equilibria.

Endogeneity of the inclusive group's identity $\boldsymbol{\theta}$, and hence of the type-specific distance costs, $\boldsymbol{h}$, plays a central role. The interactions within the inclusive group occurs with a shared identity, for instance by mixing preferred activities or by adopting a common clothing style or slang. To close the model we need to specify how $\boldsymbol{\theta}$ is determined. In line with the assumption of positive preferences for group size, we assume that $\boldsymbol{\theta}$ is chosen by the inclusive group so as to maximize the group's size (or "popularity") in the joining equilibrium that ensues,

$$\boldsymbol{\theta} = \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{j=1}^{J} \pi_j \mu_j .$$

Two factors will make the inclusive group's identity $\boldsymbol{\theta}$ align relatively closely with the endowed identity $\boldsymbol{\theta}_j$ of type $j$. First, type $j$'s *relative population share* $\pi_j$ (as a relatively large share makes type $j$ an important potential source of members for the inclusive group). Second, type $j$'s *relative aversion to cultural distance* (as a $\boldsymbol{\theta}$ close to $\boldsymbol{\theta}_j$ will encourage participation by type $j$ individuals).

**A Three-Type Case.** A three-type version of the model—i.e., one featuring natives and two immigrant groups—can be used to revisit our main empirical findings. Thus, consider the $J = 3$ case where $\Theta = [-1, 1]^2$. Let the three types be denoted by $N$, $C$, and $D$, for "natives", culturally "close" immigrants, and culturally "distant" immigrants, respectively. We assume that their endowed identities are represented by points on the unit circle: $\boldsymbol{\theta}_N = (-1, 0)$, $\boldsymbol{\theta}_C = \left(-1/2, \sqrt{3}/2\right)$ and $\boldsymbol{\theta}_D = (1, 0)$. As is illustrated below, the cultural identity of type-$C$ immigrants is then close to that of natives, whereas type-$D$ immigrants exhibit a large cultural distance from the two other types.

For comparative statics purposes, we use two parameters to represent the population type distribution: $\pi \in [0, 1]$ denotes the overall *immigrant share* and $s \in [0, 1]$ the share of culturally distant immigrants among the immigrants (henceforth, *immigrant split*). Hence $\pi_N = 1-\pi$, $\pi_C = \pi(1-s)$ and $\pi_D = \pi s$. Given this parameterization, it is natural to focus on two sets of empirical predictions from the heatmap of natives' in-group bias in Figure 5 and examine how closely the model aligns with these predictions. The first set pertains to natives' in-group bias along the line connecting the points $(NCP, MP, CP)$ in Figure 5. The corresponding comparative static exercise varies the immigrant split ($s$) while holding the immigrant share ($\pi$) constant at 40 percent. The second set of predictions concerns natives' in-group bias along the line connecting the points $(NSM, MP, OF)$ in Figure 5. The comparative statics exercise that corresponds to these empirical results considers changes in the immigrant share ($\pi$) with a constant immigrant split ($s$) of 50 percent.

Figure 13 illustrates our results. As a reference point, panel (a) zooms in on our heatmap in Figure 5 and highlights the empirical predictions for natives' in-group bias along the line $(NCP, MP, CP)$ in green and those along the line $(NSM, MP, OF)$ in blue. The remaining two panels of the figure present the corresponding comparative statics results derived from a numerical example constructed to capture the qualitative features of our empirical findings using minimal parameterization. In the example, we set $\beta = 1.1$, $h_j(d) = \gamma_j d^{\sigma}$ with a common elasticity $\sigma = 1.2$ but with type-specific constant terms: $\gamma_N = 0.3$ and $\gamma_C = \gamma_D = 0.8$. Hence, in the example, natives have a lower aversion to cultural distance than immigrants, both the culturally close and the distant immigrant type. Panel (a) shows the endogenous cultural identity of the inclusive group ($\boldsymbol{\theta}$), while panel (b) depicts natives' propensity to interact exclusively with others of their own type, that is, their non-joining rate $(1 - \mu_N)$.

First, consider the solid green lines in panels (b) and (c) of Figure 13, which vary the immigrant split ($s$) while keeping the immigrant share constant at 40 percent. When $s = 0$, the population consists of only natives and culturally close immigrants ($NCP$). Panel (b) shows that the identity of the inclusive group, $\boldsymbol{\theta}$, lies on the chord between $\boldsymbol{\theta}_C$ and $\boldsymbol{\theta}_N$, positioned closer to $\boldsymbol{\theta}_C$ (reflecting that $\gamma_C > \gamma_N$). Panel (c) reveals that when $s = 0$, about half of the natives choose to engage with immigrants within the inclusive group, while the other half choose to engage only with other natives within the native exclusive group. An increase in $s$ away from zero affects natives' behavior both directly and indirectly. The *direct* effect occurs through changes in the overall immigrant inclusive group joining rate, $(1 - s)\mu_C + s\mu_D$, holding $\boldsymbol{\theta}$ constant. If increasing $s$ raises the overall immigrant joining rate (due to $\mu_D > \mu_C$), this encourages more natives to join as well. However, this direct effect will be muted through two offsetting forces: increasing $s$ raises $\mu_C$ by shrinking the own-type exclusive group for type-$C$ immigrants, while it lowers $\mu_D$ by increasing the own-type exclusive group for type-$D$ immigrants. The *indirect* effect involves a shift in the inclusive group's identity, $\boldsymbol{\theta}$. As $s$ increases from zero towards 0.5, immigrants become increasingly fractionalized, leaving natives as the overwhelming majority type in the population. To maximize the inclusive group's size, $\boldsymbol{\theta}$ shifts closer to the native identity, reflected in the green solid line in panel (b) curving towards $\boldsymbol{\theta}_N$. Hence, as $s$ increases from zero to 0.5, natives' propensity to exclusively interact with other natives slightly decreases, and conversely their likelihood of interacting with immigrants within the inclusive group slightly increases. As the share of distant immigrants continues to grow, the inclusive group's identity, $\boldsymbol{\theta}$, shifts away from the native culture, $\boldsymbol{\theta_N}$, toward that of distant immigrants, $\boldsymbol{\theta_D}$. As a result, natives increasingly choose to interact exclusively with in-group members, and this effect is amplified by a social multiplier, so even small changes in $s$ can cause significant shifts. At the endpoint where the fraction of distant immigrants is unity ($s = 1$), natives' propensity to only interact with other natives peaks. Overall, the model predicts that natives show lower inclusive behavior under cultural polarization compared to scenarios with non-cultural or mixed polarization, reflecting our empirical findings on in-group bias along the $(NCP, MP, CP)$-line in our heatmap.

Second, consider the hatched blue lines in panels (b) and (c) of Figure 13, which vary the immigrant share ($\pi$) while keeping the immigrant split ($s$) constant at 50 percent. This exercise allows for a comparison of scenarios involving a native supermajority, mixed polarization, and overall fractionalization. As a recap, and evident in panel (a) of Figure 13, our empirical analysis revealed minor variation in in-group bias among natives across the scenarios of $(NSM, MP, OF)$. Panel (c) demonstrates that our model also predicts this pattern.[21] At low values of $\pi$, when natives are the

[21] We restrict attention in the example to $\pi \geq 0.15$ as the binary choice model with logistic errors is not a suitable

dominant majority ($NSM$), the inclusive group's identity, $\boldsymbol{\theta}$, closely aligns with that of natives, $\boldsymbol{\theta}_N$. This leads many natives to identify with immigrants by joining the inclusive group. As $\pi$ increases, two counteracting forces affect native behavior: the inclusive group's identity shifts away from native identity, discouraging engagement with immigrants, while the shrinking population share of the native type encourages engagement within the inclusive group setting. Initially, the first effect dominates, but as $\pi$ increases further, the two effects balance out. Overall, consistent with our empirical findings, the model predicts similar native behavior across scenarios of native supermajority, mixed polarization, and overall fractionalization.

When considering our two comparative static results in tandem, cultural polarization emerges as the environment least favorable for fostering inclusive behavior among natives. Thus, our framework can explain why we observe a peak in natives' in-group bias in culturally polarized classroom.[22] It is, however, worth noting that alternative economic models could be adapted to account for our findings. Indeed, our approach is closely linked to existing literature on (i) cultural identity and assimilation and (ii) the endogenous formation of friendships and homophily. Classic contributions to the former literature include Lazear (1999), Bisin and Verdier (2000), and Carvalho (2013). In the latter, Currarini et al. (2009) stands out for constructing a random matching model to investigate homophily in the creation of friendship networks, highlighting that the most significant in-group bias emerges from middle-sized groups.

## 7 Conclusion

Our lab-in-the field experiment reveals that in-group bias among natives peaks in culturally polarized classrooms, where German natives form a slim majority and Muslim immigrants a large minority. In contrast, in classrooms characterized by non-cultural polarization, mixed polarization, overall fractionalization, or a native supermajority, there are significantly lower levels of in-group bias. As mechanisms for the observed in-group bias in culturally polarized classrooms, we find that inside the classroom there are fewer intergroup friendships, less shared identity, and relatively worse treatment of natives by immigrants. Outside the classroom, these experiences translate into reduced trust and other-regarding preferences toward immigrants more broadly.

These findings provide a fresh perspective on the impact of migration-induced diversity, reconciling evidence from prior studies. It has long been argued that diversity in schools, workplaces, and neighborhoods can serve as a catalyst for close intergroup interaction, thereby helping to dismantle negative stereotypes and encourage deeper cross-cultural understanding (Allport, 1954; Pettigrew

representation in an environment where one type forms an overwhelming majority in the population.

[22] In Appendix B, we present comparative static results when there are natives and just one immigrant type. Specifically, we investigate the two limiting cases where we manipulate the immigrant share ($\pi$) with $s = 0$ and $s = 1$, respectively. In the context of our heatmaps, they correspond to traversing along a vertical ray ($s = 0$) and a horizontal ray ($s = 1$) from the origin, respectively. These comparative statics also provide an explanation for the inverted U-shape observed in-group bias when varying the immigrant share in Figure 4.

and Tropp, 2008). However, recent empirical studies have found mixed results, with migration-induced diversity either improving cross-cultural relations and understanding (Calderon et al., 2023) or failing to foster meaningful contact and resulting in hostility instead (Algan et al., 2016). What makes diversity unifying in some settings but divisive in others? Our research points to two key factors: (i) whether diversity takes the form of polarization or fractionalization and (ii) the cultural distance between groups.

From a policy perspective, our results suggest that extra efforts are needed to counteract low levels of trust in culturally polarized environments. To address this in schools, potential solutions could involve modifying the curricula to incorporate lessons focused on inclusion (Alan et al., 2021) or implementing randomized seating assignments to disrupt own-group attachment and promote cross-group friendships (Faur and Laursen, 2022). Our findings also have wider implications. For example, governments use a variety of assignment rules to allocate refugees and immigrants across regions, and researchers have recently developed algorithms to improve these practices (Bansak et al., 2018). Our findings suggest it could be important to factor in whether the destination communities will become culturally polarized.

## References


Alan, Sule, Ceren Baysan, Mert Gumren, and Elif Kubilay (2021) "Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking," *The Quarterly Journal of Economics*, 136 (4), 2147–2194.

Alesina, Alberto and Eliana La Ferrara (2000) "Participation in Heterogeneous Communities," *Quarterly Journal of Economics*, 115 (3), 847–904.

——— (2002) "Who Trusts Others?," *Journal of Public Economics*, 85 (2), 207–234.

Alesina, Alberto and Marco Tabellini (2024) "The Political Effects of Immigration: Culture or Economics?" *Journal of Economic Literature*, 62 (1), 5–46.

Algan, Yann (2018) "Trust and Social Capital," in Stiglitz, Joseph E, Jean-Paul Fitoussi, and Martine Durand eds. *For Good Measure: Advancing Research on Well-Being Metric Beyond GDP*, 283–320: Paris: OECD.

Algan, Yann, Camille Hémet, and David D Laitin (2016) "The Social Effects of Ethnic Diversity at the Local Level: A Natural Experiment with Exogenous Residential Allocation," *Journal of Political Economy*, 124 (3), 696–733.

Allport, Gordon (1954) *The Nature of Prejudice*: Cambridge, MA: Addison-Wesley Publishing.

Antecol, Heather, Ozkan Eren, and Serkan Ozbeklik (2015) "The Effect of Teacher Gender on Student Achievement in Primary School," *Journal of Labor Economics*, 33 (1), 63–89.

Balestra, Simone, Beatrix Eugster, and Helge Liebert (2022) "Peers with Special Needs: Effects and Policies," *Review of Economics and Statistics*, 104 (3), 602–618.

Bansak, Kirk, Jeremy Ferwerda, Jens Hainmueller, Andrea Dillon, Dominik Hangartner, Duncan Lawrence, and Jeremy Weinstein (2018) "Improving Refugee Integration through Data-Driven Algorithmic Assignment," *Science*, 359 (6373), 325–329.

Bazzi, Samuel, Arya Gaduh, Alexander D. Rothenberg, and Maisy Wong (2019) "Unity in Diversity? How Intergroup Contact Can Foster Nation Building," *American Economic Review*, 109 (11), 3978–4025.

Bellemare, Charles and Sabine Kröger (2007) "On Representative Social Capital," *European Economic Review*, 51 (1), 183–202.

Bifulco, Robert, Jason M. Fletcher, and Stephen L. Ross (2011) "The Effect of Classmate Characteristics on Post-secondary Outcomes: Evidence from the Add Health," *American Economic Journal: Economic Policy*, 3 (1), 25–53.

Bisin, Alberto and Thierry Verdier (2000) ""Beyond the Melting Pot": Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits," *Quarterly Journal of Economics*, 115 (3), 955–988.

Bohren, J. Aislinn, Kareem Haggag, Alex Imas, and Devin G. Pope (forthcoming) "Inaccurate Statistical Discrimination: An Identification Problem," *Review of Economics and Statistics*.

Boucher, Vincent, Semih Tumen, Michael Vlassopoulos, Jackline Wahba, and Yves Zenou (2022) "Ethnic Mixing in Early Childhood: Evidence from a Randomized Field Experiment and a Structural Model," *Unpublished Manuscript*.

Brenøe, Anne Ardila and Ulf Zölitz (2020) "Exposure to More Female Peers Widens the Gender Gap in STEM Participation," *Journal of Labor Economics*, 38 (4), 1009–1054.

Brock, William A. and Steven N. Durlauf (2001) "Discrete Choice with Social Interactions," *Review of Economic Studies*, 68 (2), 235–260.

Brown, Bradford B. (2011) "Peer Groups and Peer Cultures," in Feldman, S. S. and G. R. Elliott eds. *At the Threshold: The Developing Adolescent*, 171–196: Harvard University Press.

Calderon, Alvaro, Vasiliki Fouka, and Marco Tabellini (2023) "Racial Diversity and Racial Policy Preferences: The Great Migration and Civil Rights," *Review of Economic Studies*, 90 (1), 165–200.

Carrell, Scott E, Mark Hoekstra, and Elira Kuka (2018) "The Long-Run Effects of Disruptive Peers," *American Economic Review*, 108 (11), 3377–3415.

Carrell, Scott E, Mark Hoekstra, and James E West (2019) "The Impact of College Diversity on

Behavior Toward Minorities," *American Economic Journal: Economic Policy*, 11 (4), 159–182.

Carrell, Scott E and James E West (2010) "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors," *Journal of Political Economy*, 118 (3), 409–432.

Carrell, Scott and Mark Hoekstra, "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids," *American Economic Journal: Applied Economics*, 2 (1), 2010.

Carvalho, Jean-Paul (2013) "Veiling," *Quarterly Journal of Economics*, 128 (1), 337–370.

Chetty, Raj, John Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan (2011) " How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star," *Quarterly Journal of Economics*, 126 (4), 1593–1660.

Chuard, Caroline, Annatina Aerne, Simone Balestra, Beatrix Eugster, and Roland Hodler (2022) "Ethnic Clustering in Schools and Early Career Outcomes," CEPR Discussion Papers 17505, C.E.P.R. Discussion Papers.

Corno, Lucia, Eliana La Ferrara, and Justine Burns (2022) "Interaction, Stereotypes, and Performance: Evidence from South Africa," *American Economic Review*, 112 (12), 3848–75.

Currarini, Sergio, Matthew O Jackson, and Paolo Pin (2009) "An Economic Model of Friendship: Homophily, Minorities, and Segregation," *Econometrica*, 77 (4), 1003–1045.

Dahl, Gordon B, Christina Felfe, Paul Frijters, and Helmut Rainer (2022) "Caught between Cultures: Unintended Consequences of Improving Opportunity for Immigrant Girls," *Review of Economic Studies*, 89 (5), 2491–2528.

Dahl, Gordon, Andreas Kotsadam, and Dan-Olof Rooth (2021) "Does Integration Change Gender Attitudes? The Effect of Randomly Assigning Women to Traditionally Male Teams," *Quarterly Journal of Economics*, 136 (2), 987–1030.

Dee, Thomas S. (2004) "Are There Civic Returns to Education?" *Journal of Public Economics*, 88 (9), 1697–1720.

Depetris-Chauvin, Emilio, Ruben Durante, and Filipe Campante (2020) "Building Nations through Shared Experiences: Evidence from African Football," *American Economic Review*, 110 (5), 1572–1602.

Egorov, Georgy, Ruben Enikolopov, Alexey Makarin, and Maria Petrova (2021) "Divided we stay home: Social distancing and ethnic diversity," *Journal of Public Economics*, 194, 104328.

Falk, Armin and Christian Zehnder (2013) "A City-Wide Experiment on Trust Discrimination," *Journal of Public Economics*, 100, 15–27.

Faur, Sharon and Brett Laursen (2022) "Classroom Seat Proximity Predicts Friendship Formation," *Frontiers in Psychology*, 13, 796002.

Federal Anti-Discrimination Agency (2022) *Diskriminierung an Schulen erkennen und vermeiden: Praxileitfaden zum Abbau von Diskriminierung an Schulen*: Antidiskrimierungsstelle des Bundes.

Federal Government of North Rhine-Westphalia (2022) *11. Kinder-und Jugendbericht der Landesregierung Nordrhein-Westfalen*: Ministerium für Kinder, Familie, Flüchtlinge und Integration des Landes Nordrhein-Westfalen.

Federal Statistical Office of Germany (2009) *Bevölkerung und Erwerbstätigkeit. Bevölkerung mit Migrationshintergrund - Ergebnisse des Mikrozensus 2018*: Statistische Bundesamt.

Feld, Jan and Ulf Zölitz (2017) "Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects," *Journal of Labor Economics*, 35 (2), 387–428.

Felfe, Christina, Martin G Kocher, Helmut Rainer, Judith Saurer, and Thomas Siedler (2021) "More Opportunity, More Cooperation? The Behavioral Effects of Birthright Citizenship on Immigrant Youth," *Journal of Public Economics*, 200, 104448.

Fershtman, Chaim and Uri Gneezy (2001) "Discrimination in a Segmented Society: An Experimental Approach," *Quarterly Journal of Economics*, 116 (1), 351–377.

Figlio, David N, Paola Giuliano, Riccardo Marchingiglio, Umut Ãzek, and Paola Sapienza (2024) "Diversity in Schools: Immigrants and the Educational Performance of U.S. Born Students," *Review of Economic Studies*, 91 (2), 972–1006.

Hanson, Gordon and Craig McIntosh (2016) "Is the Mediterranean the New Rio Grande? US and EU Immigration Pressures in the Long Run," *Journal of Economic Perspectives*, 30 (4), 57–82.

Hanushek, Eric, John Kain, Jacob Markman, and Steven Rivkin (2003) "Does Peer Ability Affect Student Achievement?," *Journal of Applied Econometrics*, 18 (5), 527–544.

Horowitz, Donald L (1985) *Ethnic Groups in Conflict*: University of California Press.

Hotelling, Harold (1929) "Stability in Competition," *Economic Journal*, 39 (153), 41–57.

Hoxby, Caroline (2002) "The Power of Peers: How Does the Makeup of a Classroom Influence Achievement," *Education Next*, 20 (2), 56–63.

Kremer, Michael and Dan Levy (2008) "Peer Effects and Alcohol Use among College Students," *Journal of Economic Perspectives*, 22 (3), 189–206.

Kruse, Mathias (2023) "Closing the Solidarity Gap? How Ethnic Diversity Alters Who We Are Willing to Support," *Political Psychology*, 44 (3), 531–549.

Lavy, Victor, M Daniele Paserman, and Analia Schlosser (2012) "Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom," *Economic Journal*, 122 (559), 208–237.

Lazear, Edward P (1999) "Culture and Language," *Journal of Political Economy*, 107 (S6), S95–S126.

Lowe, Matt (2021) "Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration," *American Economic Review*, 111 (6), 1807–1844.

Luttmer, Erzo FP (2001) "Group Loyalty and the Taste for Redistribution," *Journal of Political Economy*, 109 (3), 500–528.

Lyle, David (2007) "Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point," *Review of Economics and Statistics*, 89 (2), 289–299.

Manski, Charles (1993) "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60 (31), 531–542.

McAuliffe, M. and A. Triandafyllidou (2021) *World Migration Report 2022*: United Nations International Organization for Migration, Geneva.

Milligan, Kevin, Enrico Moretti, and Philip Oreopoulos (2004) "Does Education Improve Citizenship? Evidence from the United States and the United Kingdom," *Journal of Public Economics*, 88 (9), 1667–1695.

Montalvo, José G and Marta Reynal-Querol (2005) "Ethnic Polarization, Potential Conflict, and Civil Wars," *American Economic Review*, 95 (3), 796–816.

Mousa, Salma (2020) "Building Social Cohesion between Christians and Muslims through Soccer in Post-ISIS Iraq," *Science*, 369 (6505), 866–870.

Nieden, Birgit zur and Juliane Karakayali (2016) "Harte Tür: Schulische Segregation nach Herkunft in der postmigrantischen Gesellschaft," 81–96.

Ohinata, Asako and Jan C Van Ours (2013) "How Immigrant Children Affect the Academic Achievement of Native Dutch Children," *Economic Journal*, 123 (570), F308–F331.

Pettigrew, Thomas F and Linda R Tropp (2008) "How Does Intergroup Contact Reduce Prejudice? Meta-Analytic Tests of Three Mediators," *European Journal of Social Psychology*, 38 (6), 922–934.

Puskeppeleit, Jürgen and Marianne Krüger-Potratz (1999) *Bildungspolitik und Migration: Texte und Dokumente zur Beschulung ausländischer und ausgesiedelter Kinder und Jugendlicher: 1950-1999*: Bd 1. Arbeitsstelle Interkulturelle Pädagogik.

Rao, Gautam (2019) "Familiarity Does Not Breed Contempt: Generosity, Discrimination, and Diversity in Delhi Schools," *American Economic Review*, 109 (3), 774–809.

Sacerdote, Bruce (2001) "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *Quarterly Journal of Economics*, 116 (2), 681–704.

——— (2011) "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" in Hanushek, Eric A., Stephen Machin, and Ludger Woessmann eds. *Handbook of the Economics of Education*, 3, 249–277: Elsevier.

Sapienza, Paola, Anna Toldra-Simats, and Luigi Zingales (2013) "Understanding trust," *The Economic Journal*, 123 (573), 1313–1332.

Tropp, Linda R and Suchi Saxena (2018) “Re-Weaving the Social Fabric Through Integrated Schools: How Intergroup Contact Prepares Youth to Thrive in a Multiracial Society,” *Research Brief No. 13., National Coalition on School Diversity*.

Wells, Amy Stuart, Lauren Fox, and Diana Cordova-Cobo (2016) “How Racially Diverse Schools and Classrooms Can Benefit all Students,” *The Education Digest*, 82 (1), 17.

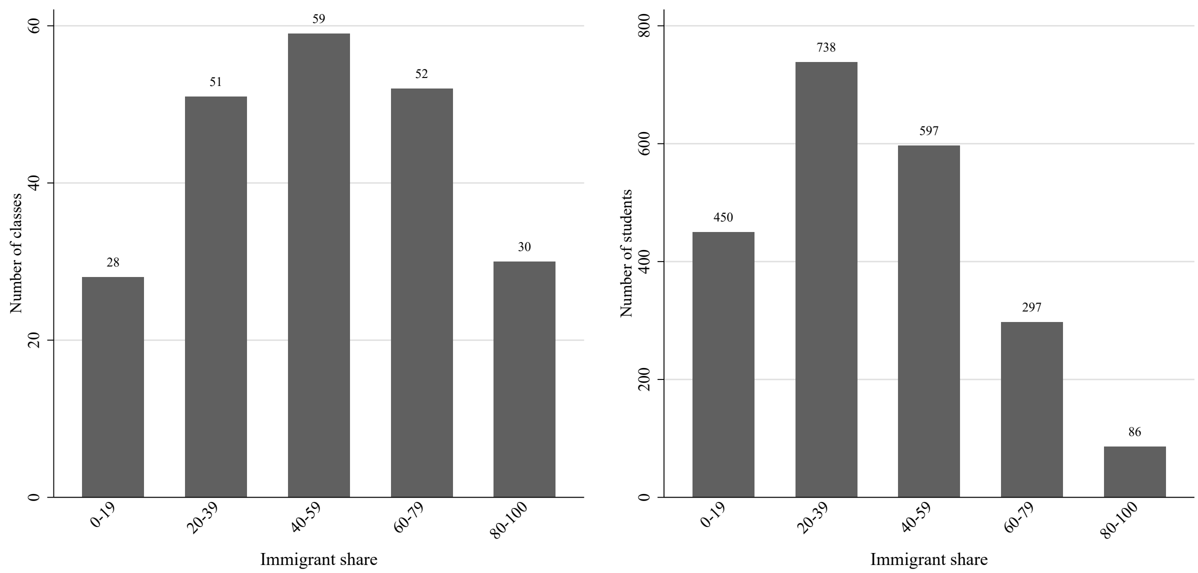



**Figure 1.** Share of Immigrant Peers at the Classroom and Student Level

*Notes:* There are 220 classes in the left panel and 2,168 native German students in the right panel.

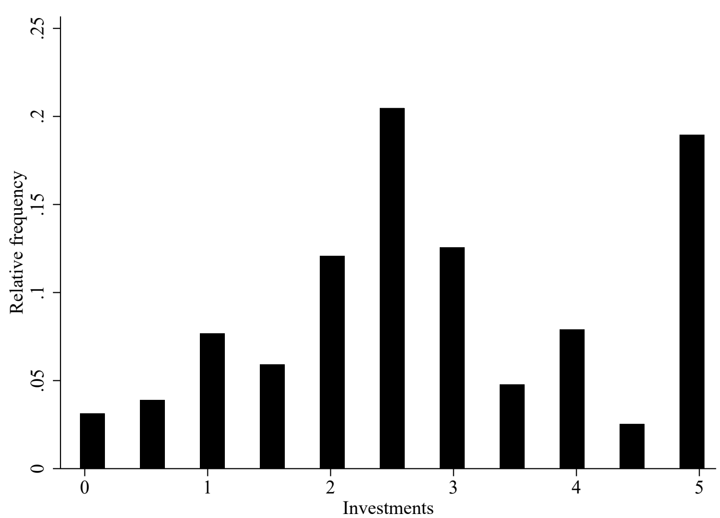



**Figure 2.** Investment Decisions Made by Native Senders

*Notes:* Histogram of investment decisions made by native senders. Senders could invest between 0 and 5 euros, in 50 cent increments.

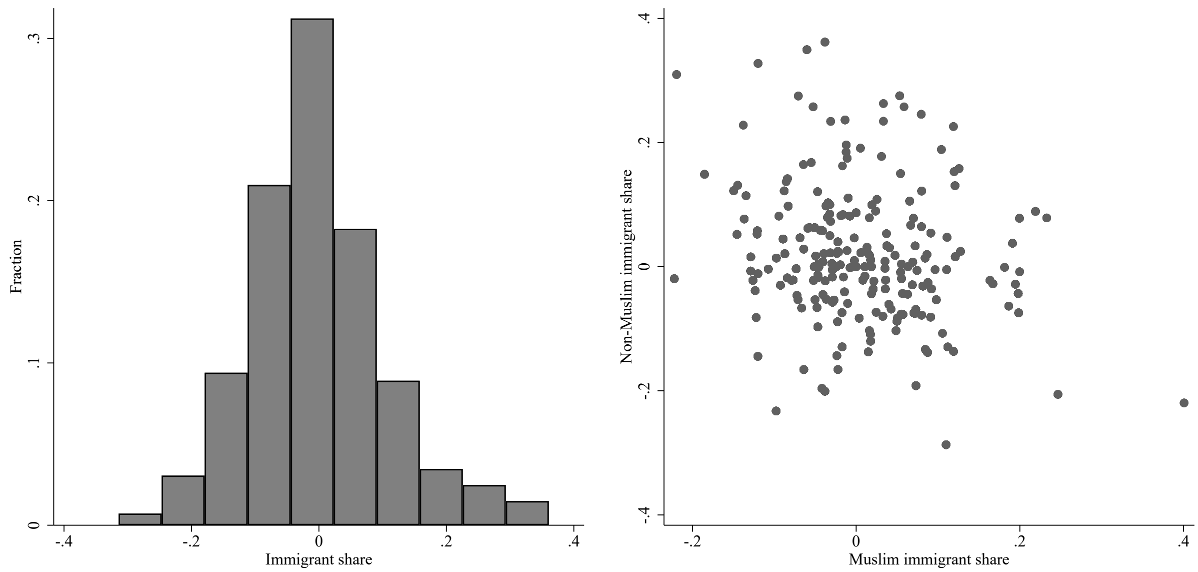



**Figure 3.** Identifying Variation after Removing School Fixed Effects

*Notes:* The left panel is a histogram of the residualized immigrant share, net of school fixed effects. The right panel is a scatter plot of the residualized non-Muslim immigrant share against the residualized Muslim immigrant share, both net of school fixed effects.

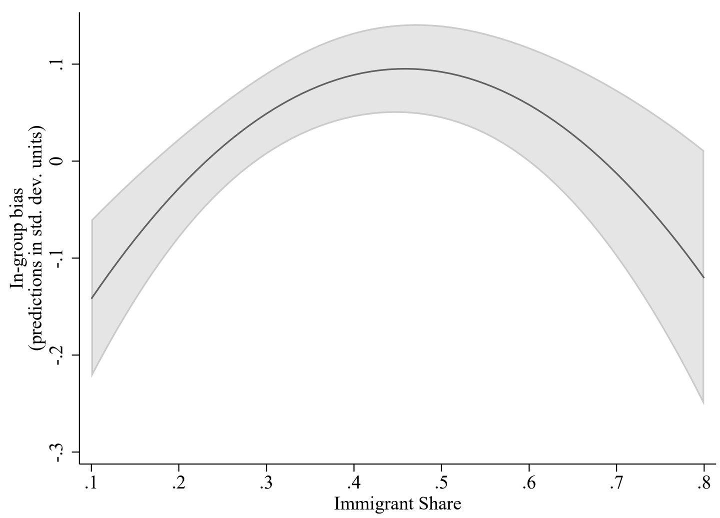
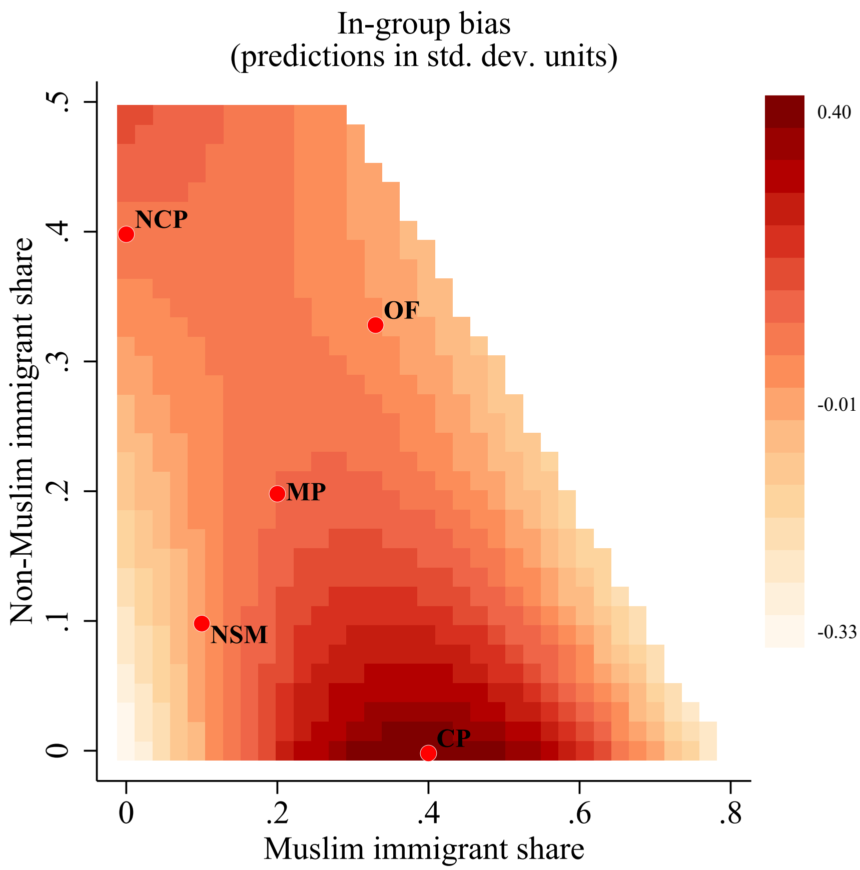



**Figure 4.** In-Group Bias: An Inverted-U in the Share of Immigrant Peers

*Notes:* The figure uses estimates from column 3 of Table 3 to predict the in-group out-group investment gap for immigrant shares, evaluating control variables at their means. The grey shaded area denotes pointwise 90% confidence intervals.
N=2,168. 1 standard deviation = 76 euro-cents.




| | Differences between classroom types (in std. dev. units) |
|---|---|
| CP-NCP | 0.333** |
| | (0.156) |
| CP-MP | 0.318** |
| | (0.147) |
| CP-OF | 0.404** |
| | (0.168) |
| CP-NSM | 0.424*** |
| | (0.137) |

**Figure 5.** In-Group Bias: How Type of Classroom Diversity Matters

*Notes:* In-group bias is measured by the in-group out-group investment gap in the first stage of the investment game. The heatmap shows predicted values for in-group bias based on estimates for equation (5). The peak in in-group bias occurs in a classroom exhibiting **cultural polarization** (**CP**), with native German peers forming a slight majority group [60%] and Muslim immigrant peers a large minority group [40%]. The table shows the difference between predicted in-group bias at this peak and four other classroom scenarios:

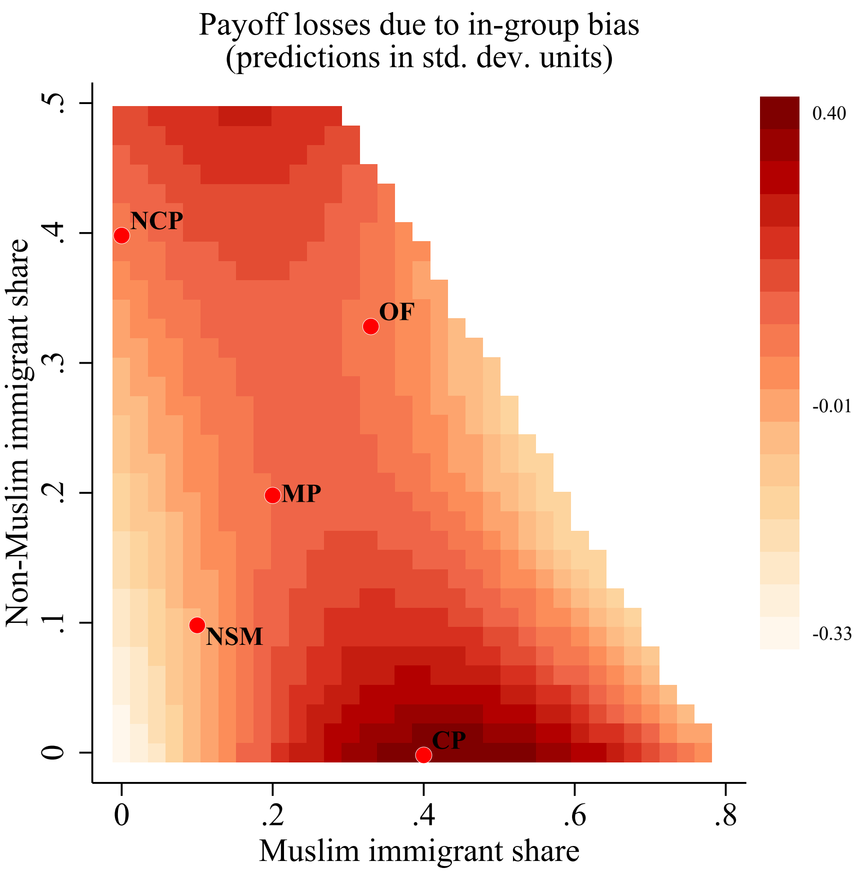
**Non-cultural polarization (NCP)**: native German peers form a slight majority group [60%] and non-Muslim immigrant peers a large minority group [40%]

**Mixed polarization (MP)**: native peers form a slight majority group [60%] and both Muslim and non-Muslim immigrant peers equally large, medium-sized minority groups [20% each]

**Overall fractionalization (OF)**: native peers, Muslim immigrant peers, and non-Muslim immigrant peers form equally large groups [33% each]

**Native supermajority (NSM)**: native peers form a large majority group [80%] and both Muslim and non-Muslim immigrant peers small minority groups [10% each]

N=2,168. 1 standard deviation = 76 euro-cents. *** p<0.01, ** p<0.05, * p<0.1.




| Differences between classroom types (in std. dev. units) | |
|---|---|
| CP-NCP | 0.344** |
| | (0.172) |
| CP-MP | 0.302* |
| | (0.165) |
| CP-OF | 0.331* |
| | (0.193) |
| CP-NSM | 0.442*** |
| | (0.157) |

**Figure 6.** Payoff Losses Due to In-Group Bias

*Notes:* Payoff losses due to in-group bias are the differences in expected payoffs of senders when randomly matched with an immigrant versus a native interaction partner. The heatmap shows predicted values for payoff losses based on estimates for equation (5). The table shows the difference between predicted payoff losses in a classroom exhibiting **cultural polarization** (**CP**) [60% natives, 40% Muslim immigrants] and four other classroom scenarios:

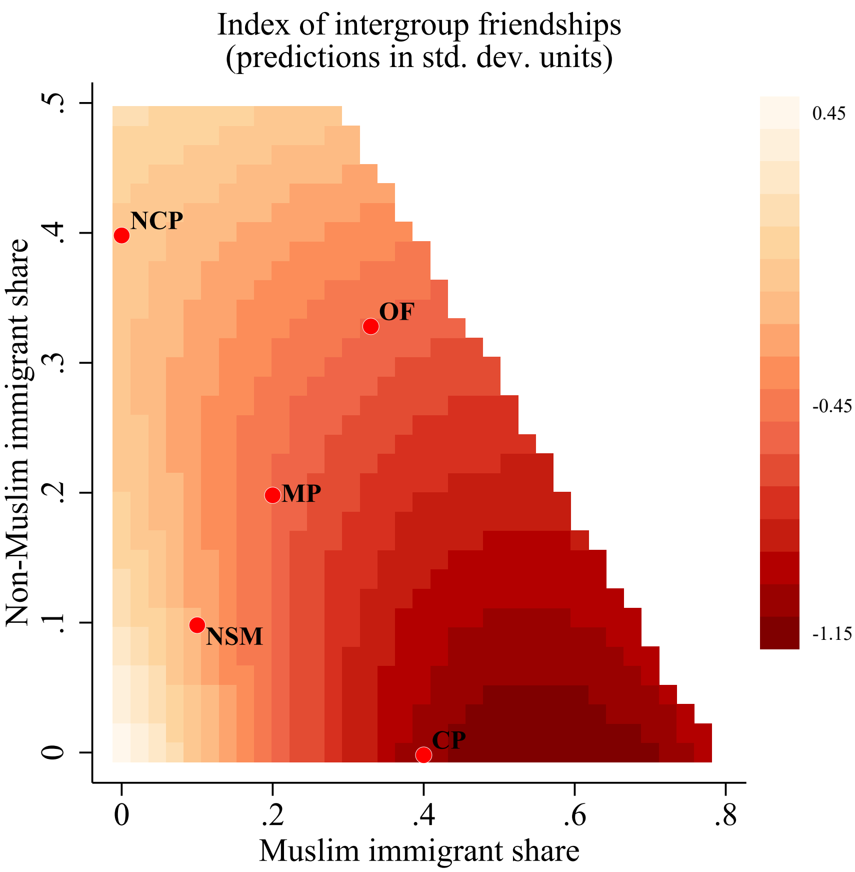
**Non-cultural polarization (NCP)**: [60% natives, 40% non-Muslim immigrants];

**Mixed polarization (MP)**: [60% natives, 20% Muslim immigrants, 20% non-Muslim immigrants];

**Overall fractionalization (OF)**: [34% natives, 33% Muslim immigrants, 33% non-Muslim immigrants];

**Native supermajority (NSM)**:[80% natives, 10% Muslim immigrants, 10% non-Muslim immigrants].

N=2,168. 1 standard deviation = 36 euro-cents. *** p<0.01, ** p<0.05, * p<0.1.




| Differences between classroom types (in std. dev. units) | |
|---|---|
| CP-NCP | -1.036*** |
| | (0.277) |
| CP-MP | -0.540** |
| | (0.263) |
| CP-OF | -0.548* |
| | (0.320) |
| CP-NSM | -0.852*** |
| | (0.258) |

**Figure 7.** Intergroup Friendships

*Notes:* The intergroup friendship index measures the degree of potential overlap between the friendships of natives and immigrants within a class (see the text for a detailed description). The heatmap shows predicted values for the intergroup friendship based on estimates for equation (5). The table shows the difference between predicted payoff losses in a classroom exhibiting **cultural polarization** (**CP**) [60% natives, 40% Muslim immigrants] and four other classroom scenarios:

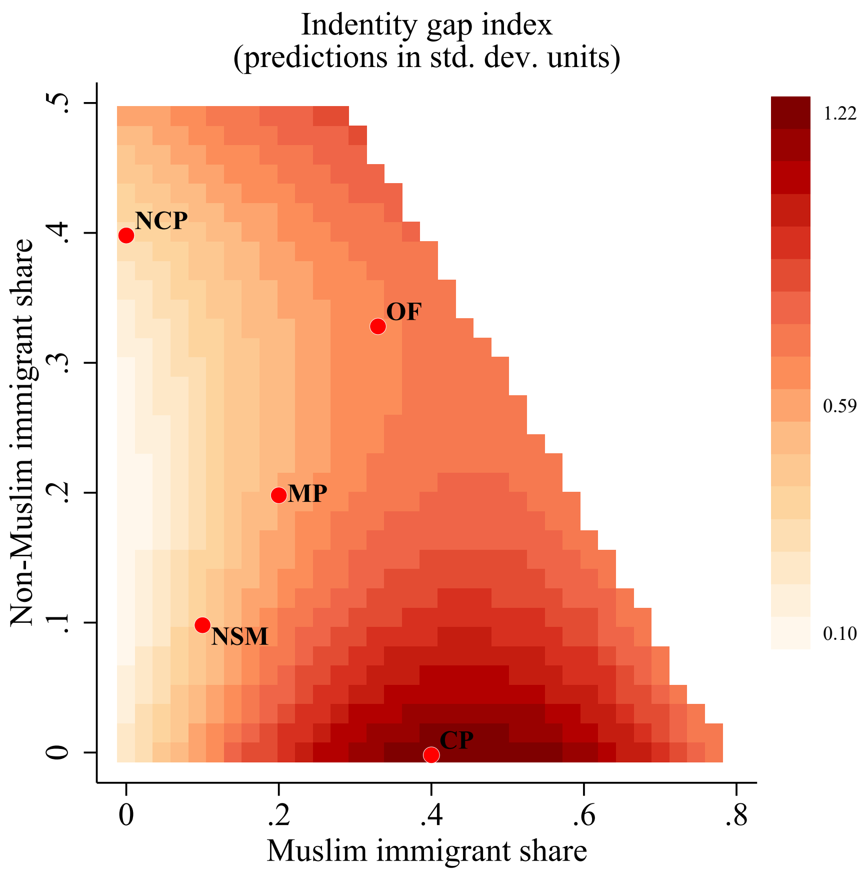
**Non-cultural polarization (NCP)**: [60% natives, 40% non-Muslim immigrants];

**Mixed polarization (MP)**: [60% natives, 20% Muslim immigrants, 20% non-Muslim immigrants];

**Overall fractionalization (OF)**: [34% natives, 33% Muslim immigrants, 33% non-Muslim immigrants];

**Native supermajority (NSM)**:[80% natives, 10% Muslim immigrants, 10% non-Muslim immigrants].

N=2,176. 1 standard deviation = 5.20 percentage points. *** p<0.01, ** p<0.05, * p<0.1.




| Differences between classroom types (in std. dev. units) | |
|---|---|
| CP-NCP | 0.942*** |
| | (0.294) |
| CP-MP | 0.691** |
| | (0.280) |
| CP-OF | 0.560* |
| | (0.336) |
| CP-NSM | 0.814*** |
| | (0.295) |

**Figure 8.** Shared Identity

*Notes:* The identiy gap index measures the difference between each native student's level of self-identification as German and the average level of self-identification among their immigrant classmates The heatmap shows predicted values for the intergroup friendship based on estimates for equation (5). The table shows the difference between predicted payoff losses in a classroom exhibiting **cultural polarization** (**CP**) [60% natives, 40% Muslim immigrants] and four other classroom scenarios:

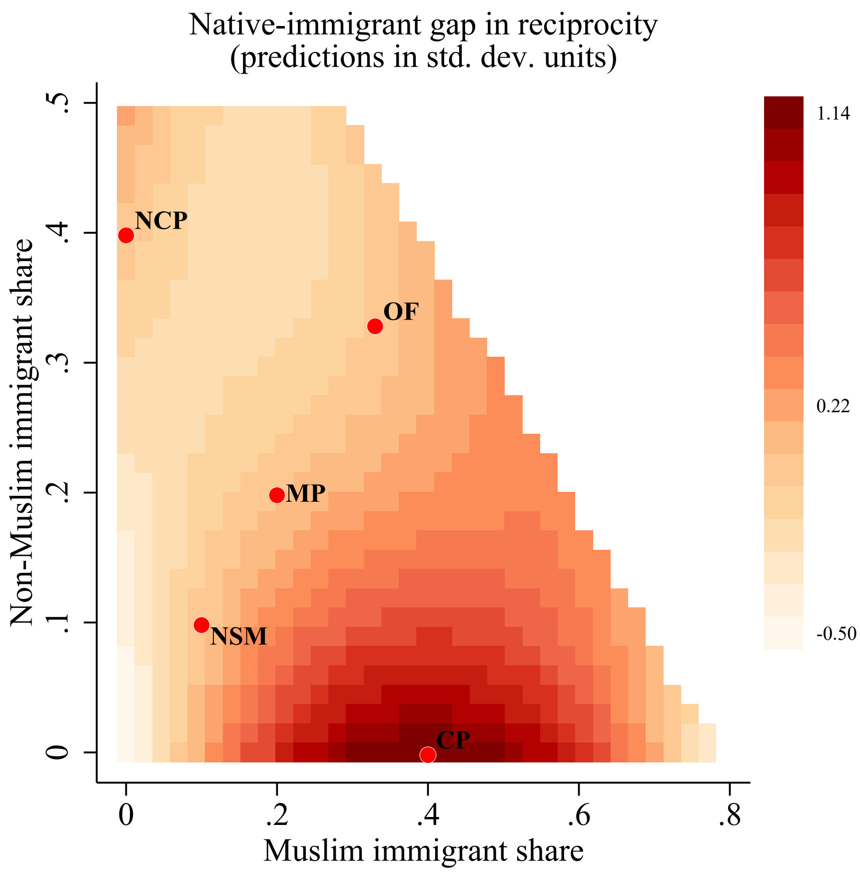
**Non-cultural polarization (NCP)**: [60% natives, 40% non-Muslim immigrants];

**Mixed polarization (MP)**: [60% natives, 20% Muslim immigrants, 20% non-Muslim immigrants];

**Overall fractionalization (OF)**: [34% natives, 33% Muslim immigrants, 33% non-Muslim immigrants];

**Native supermajority (NSM)**:[80% natives, 10% Muslim immigrants, 10% non-Muslim immigrants].

N=2,195. 1 standard deviation = 0.99 points on a 1-4 agreement scale. *** p<0.01, ** p<0.05, * p<0.1




| Differences between classroom types (in std. dev. units) | |
|---|---|
| CP-NCP | 1.166** |
| | (0.531) |
| CP-MP | 1.138** |
| | (0.543) |
| CP-OF | 1.207* |
| | (0.671) |
| CP-NSM | 1.185** |
| | (0.575) |

**Figure 9.** Native vs. Immigrant Classmates' Behavior as Receiver

*Notes:* For each individual $i$, the native-immigrant gap in classmates' reciprocity is the difference between how much, on average, their native and immigrant classmates transfer back to natives in the second stage of the investment game (with the back transfers of each classmate averaged over the 11 possible investments from a native sender). The heatmap shows predicted values for this gap based on estimates for equation (5). The table shows the difference between the predicted gap in native vs. immigrant classmates' return behavior in a classroom exibiting **cultural polarization** (**CP**) [60% natives, 40% Muslim immigrants] and four other classroom scenarios:

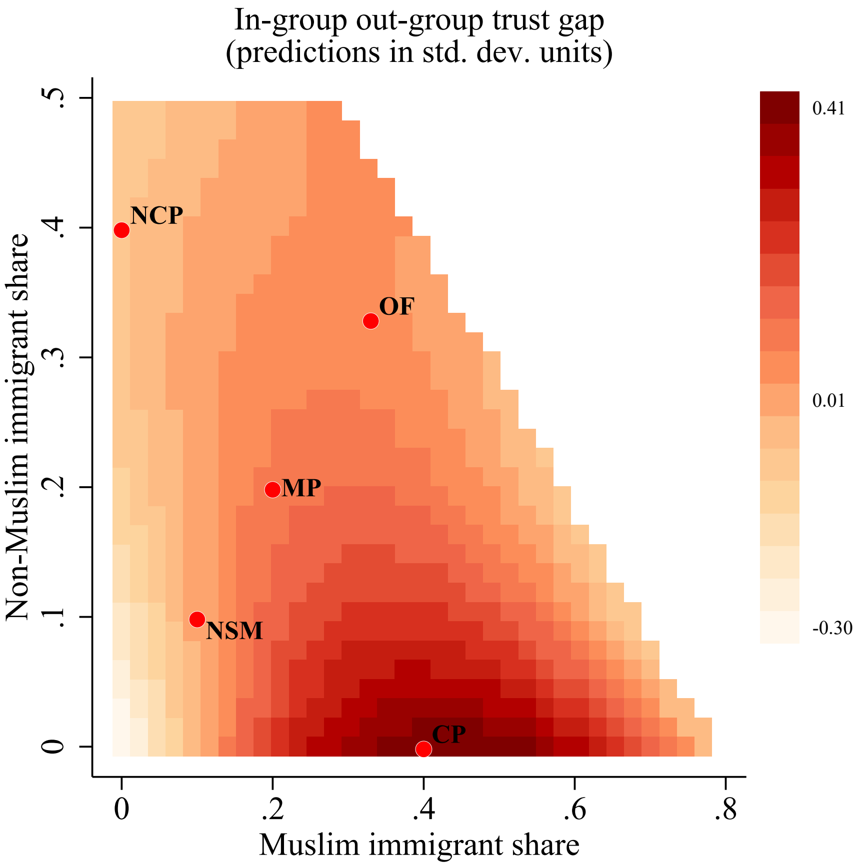
**Non-cultural polarization (NCP)**: [60% natives, 40% non-Muslim immigrants];

**Mixed polarization (MP)**: [60% natives, 20% Muslim immigrants, 20% non-Muslim immigrants];

**Overall fractionalization (OF)**: [34% natives, 33% Muslim immigrants, 33% non-Muslim immigrants];

**Native supermajority (NSM)**:[80% natives, 10% Muslim immigrants, 10% non-Muslim immigrants].

N=2,105. 1 standard deviation = 1.28 euros. *** p<0.01, ** p<0.05, * p<0.1




| Differences between classroom types (in std. dev. units) | |
|---|---|
| CP-NCP | 0.495** |
| | (0.196) |
| CP-MP | 0.333* |
| | (0.189) |
| CP-OF | 0.382* |
| | (0.225) |
| CP-NSM | 0.426** |
| | (0.176) |

**Figure 10.** Biased Beliefs about Immigrants' Trustworthiness

*Notes:* The in-group out-group trust gap is measured through survey questions asking students how much they trust people with German and foreign nationality, respectively. The heatmap shows predicted values for the trust gap based on estimates for equation (5). The table shows the difference between predicted in-group out-group trust gap in a classroom exibiting **cultural polarization** (**CP**) [60% natives, 40% Muslim immigrants] and four other classroom scenarios:

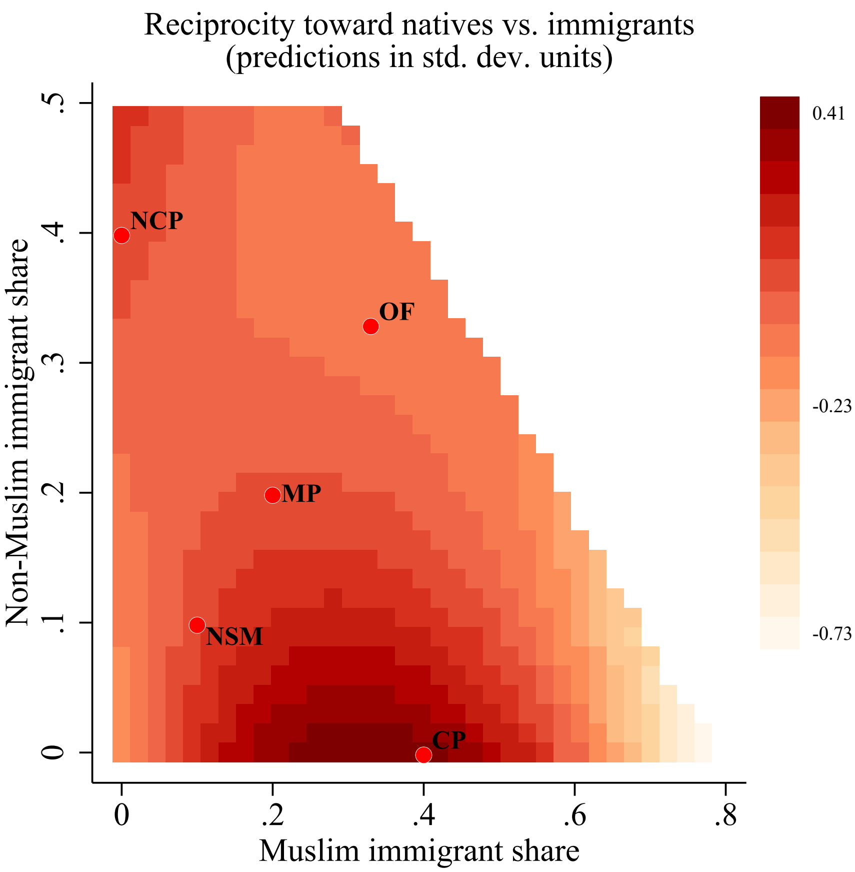
**Non-cultural polarization (NCP)**: [60% natives, 40% non-Muslim immigrants];

**Mixed polarization (MP)**: [60% natives, 20% Muslim immigrants, 20% non-Muslim immigrants];

**Overall fractionalization (OF)**: [34% natives, 33% Muslim immigrants, 33% non-Muslim immigrants];

**Native supermajority (NSM)**:[80% natives, 10% Muslim immigrants, 10% non-Muslim immigrants].

N=2,113. 1 standard deviation = 1.59 points on a 0-10 Lickert scale. *** p<0.01, ** p<0.05, * p<0.1.




| Differences between classroom types (in std. dev. units) | |
|---|---|
| CP-NCP | 0.341* |
| | (0.179) |
| CP-MP | 0.371* |
| | (0.190) |
| CP-OF | 0.487** |
| | (0.218) |
| CP-NSM | 0.346** |
| | (0.175) |

**Figure 11.** In-Group Biased Social Preferences: Natives' Behavior when Playing as the Receiver
*Notes:* Reciprocity toward natives versus immigrants is measured by how much, on average, native receivers transfer back to native versus immigrant senders in the second stage of the investment game. The heatmap shows predicted values for the in-group out-group return gap based on estimates for equation (5). The table shows the difference between the predicted in-group out-group return gap in a classroom exibiting **cultural polarization** (**CP**) [60% natives, 40% Muslim immigrants] and four other classroom scenarios:

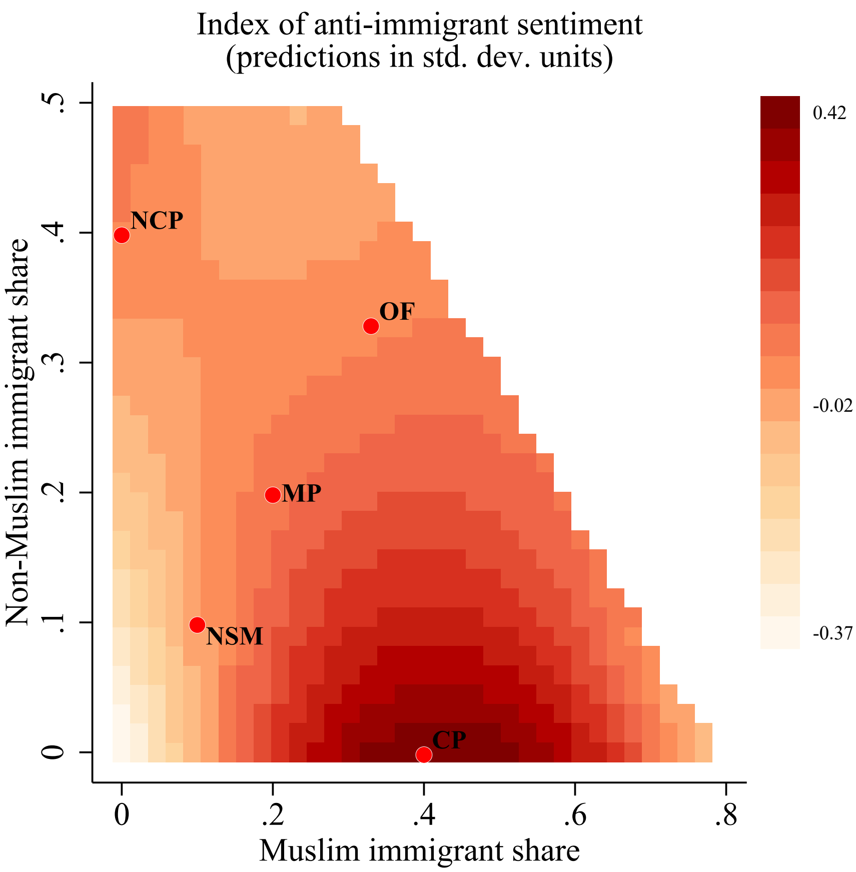
**Non-cultural polarization (NCP)**: [60% natives, 40% non-Muslim immigrants];

**Mixed polarization (MP)**: [60% natives, 20% Muslim immigrants, 20% non-Muslim immigrants];

**Overall fractionalization (OF)**: [34% natives, 33% Muslim immigrants, 33% non-Muslim immigrants];

**Native supermajority (NSM)**:[80% natives, 10% Muslim immigrants, 10% non-Muslim immigrants].

N=2,093. 1 standard deviation = 1.05 euros. *** p<0.01, ** p<0.05, * p<0.1.




| Differences between classroom types (in std. dev. units) | |
|---|---|
| CP-NCP | 0.397* |
| | (0.202) |
| CP-MP | 0.362** |
| | (0.179) |
| CP-OF | 0.407* |
| | (0.213) |
| CP-NSM | 0.470*** |
| | (0.177) |

**Figure 12.** Anti-Immigrant Sentiment

*Notes:* Anti-immigration sentiment is measured by students' disagreement with three survey questions asking whether it is fair that workers of Turkish, Polish, and French descent are allowed to work in Germany (4=strongly disagree, 3=disagree somewhat, 2=agree somewhat, 1=strongly agree). Using a principal component analysis, we create an index of anti-immigration sentiment based on these questions. We normalize the index to have mean 0 and standard deviation 1. The heatmap shows predicted values for the index of anti-immigration sentiment based on estimates for equation (5). The table shows the difference between predicted values of anti-immigration sentiment in a classroom exibiting **cultural polarization** (**P1**) [60% natives, 40% Muslim immigrants] and four other classroom scenarios:

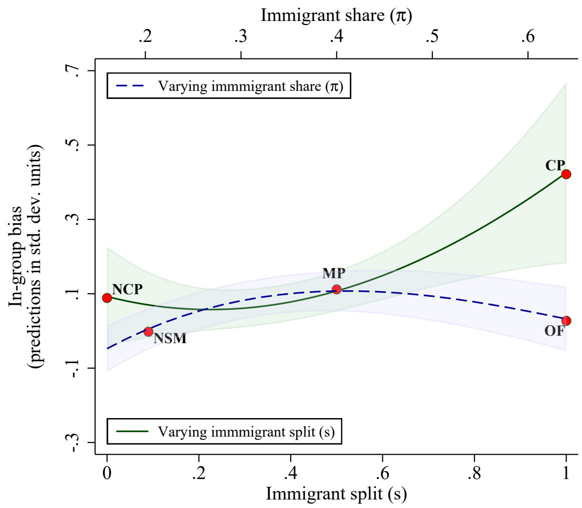
**Non-cultural polarization (NCP)**: [60% natives, 40% non-Muslim immigrants];

**Mixed polarization (MP)**: [60% natives, 20% Muslim immigrants, 20% non-Muslim immigrants];

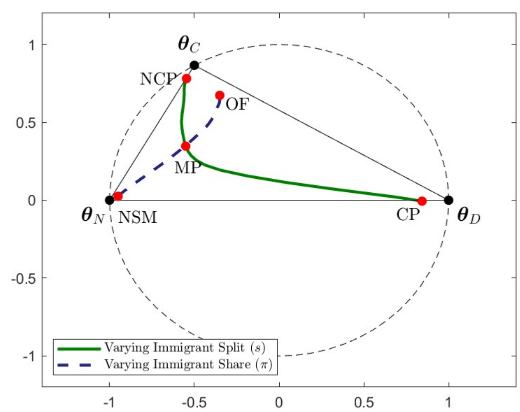
**Overall fractionalization (OF)**: [34% natives, 33% Muslim immigrants, 33% non-Muslim immigrants];

**Native supermajority (NSM)**:[80% natives, 10% Muslim immigrants, 10% non-Muslim immigrants].

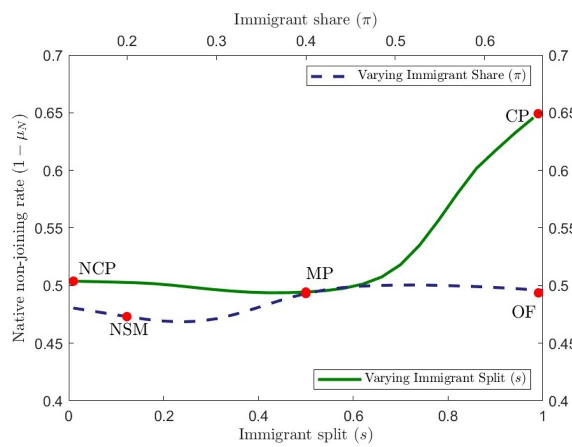
N=2,095. 1 standard deviation = 1.64. *** p<0.01, ** p<0.05, * p<0.1.




**(a)** Empirical Predictions for In-group Bias




**(b)** Cultural Identity $\boldsymbol{\theta}$ of the Inclusive Group




**(c)** In-Group Bias: Native Own Group Non-Joining Rate $1-\mu_N$

**Figure 13.** Comparative Statics of the Model with Natives and Two Immigrant Types

**Table 1.** Summary Statistics

| | Mean | Std. Dev. |
|---|---|---|
| *Main dependent variable* | | |
| In-group/out-group investment gap | 0.094 | 0.758 |
| *Main independent variables* | | |
| Immigrant share | 0.391 | 0.217 |
| Muslim immigrant share | 0.190 | 0.180 |
| Non-Muslim immigrant share | 0.201 | 0.112 |
| *Background variables* | | |
| Male | 0.540 | 0.499 |
| Age | 15.828 | 0.620 |
| Age missing | 0.017 | 0.130 |
| Catholic | 0.138 | 0.345 |
| Protestant | 0.531 | 0.499 |
| Muslim | 0.012 | 0.109 |
| Other religion | 0.318 | 0.466 |
| SES: Two-parent hh; high education | 0.199 | 0.400 |
| SES: Two-parent hh; low education | 0.268 | 0.443 |
| SES: Single-parent hh; high education | 0.070 | 0.255 |
| SES: Single-parent hh; low education | 0.265 | 0.442 |
| SES: missing | 0.198 | 0.398 |
| Age mother | 46.169 | 8.483 |
| Age mother: missing | 0.052 | 0.221 |
| Age father | 50.306 | 11.075 |
| Age father: missing | 0.087 | 0.281 |
| Observations | 2,168 | |

*Notes*: Summary statistics for the main estimation sample. When defining Socioeconomic status (SES), a high education household (hh) is one where at least one parent has either a high school or university degree.

**Table 2.** Balancing Tests with and without School Fixed Effects

| Dependent variable: | Immigrant share | | Muslim Immigrant share | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Male | -0.005 | -0.005 | -0.007 | -0.004 |
| | (0.010) | (0.005) | (0.008) | (0.003) |
| Age | 0.063*** | -0.003 | 0.051*** | 0.002 |
| | (0.012) | (0.005) | (0.009) | (0.004) |
| Age missing | -0.038 | -0.030 | -0.026 | -0.018 |
| | (0.035) | (0.019) | (0.030) | (0.015) |
| Protestant | -0.087*** | 0.010 | -0.088*** | 0.002 |
| | (0.020) | (0.008) | (0.016) | (0.007) |
| Muslim | 0.123*** | -0.008 | 0.120** | 0.005 |
| | (0.044) | (0.021) | (0.050) | (0.022) |
| Other or no religion | -0.100*** | 0.005 | -0.104*** | -0.004 |
| | (0.022) | (0.009) | (0.018) | (0.007) |
| SES: Two-parent hh; low education | 0.041*** | -0.005 | 0.034*** | 0.002 |
| | (0.014) | (0.007) | (0.010) | (0.004) |
| SES: Single-parent hh; high education | 0.023 | 0.001 | 0.018 | 0.004 |
| | (0.018) | (0.009) | (0.012) | (0.006) |
| SES: Single-parent hh; low education | 0.062*** | -0.003 | 0.053*** | 0.007 |
| | (0.014) | (0.006) | (0.012) | (0.005) |
| SES: missing | 0.085*** | 0.005 | 0.077*** | 0.012* |
| | (0.015) | (0.007) | (0.013) | (0.007) |
| Age mother | -0.002* | 0.001 | -0.002** | 0.000 |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Age mother: missing | 0.076* | -0.012 | 0.087** | 0.008 |
| | (0.045) | (0.022) | (0.042) | (0.022) |
| Age father | -0.001 | -0.000 | -0.002* | -0.000 |
| | (0.001) | (0.000) | (0.001) | (0.000) |
| Age father: missing | 0.059 | 0.020 | 0.069* | 0.028 |
| | (0.045) | (0.019) | (0.039) | (0.019) |
| Class size | -0.001 | 0.002 | -0.004* | -0.002 |
| | (0.003) | (0.003) | (0.002) | (0.002) |
| Observations | 2,168 | 2,168 | 2,168 | 2,168 |
| R-squared | 0.11 | 0.75 | 0.15 | 0.77 |
| F-statistic | 7.70 | 1.13 | 8.34 | 1.05 |
| p-value | 0.000 | 0.329 | 0.000 | 0.409 |

*Notes:* Columns 1 and 2 report OLS regressions with the fraction of immigrant peers as the dependent variable, while columns 3 and 4 use the fraction of Muslim immigrant peers as the dependent variable. Standard errors are reported in parentheses and are clustered at the classroom level. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

**Table 3.** In-Group Bias: An Inverted-U in the Share of Immigrant Peers

| Dependent variable: | in-group out-group investment gap | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Panel A: Linear Specification | | | |
| Immigrant share | 0.102 | 0.0828 | 0.0957 |
| | (0.162) | (0.165) | (0.163) |
| Observations | 2,168 | 2,168 | 2,168 |
| R-squared | 0.049 | 0.055 | 0.060 |
| Panel B: Quadratic Specification | | | |
| Immigrant share | 1.642*** | 1.706*** | 1.693*** |
| | (0.396) | (0.397) | (0.395) |
| Immigrant share squared | -1.779*** | -1.876*** | -1.847*** |
| | (0.452) | (0.455) | (0.451) |
| p-value: coeffs. jointly equal to zero | 0.0001 | 0.0001 | 0.0001 |
| p-value: coeffs. equal but opposite in sign | 0.415 | 0.319 | 0.359 |
| Observations | 2,168 | 2,168 | 2,168 |
| R-squared | 0.054 | 0.060 | 0.064 |
| Basic controls | ✓ | ✓ | ✓ |
| Religious background | | ✓ | ✓ |
| Family background | | | ✓ |

*Notes:* OLS estimates of equation (3). The dependent variable is the in-group out-group investment gap, normalized to be mean 0 and standard deviation 1. Basic controls include school fixed effects, a student's gender and age, and class size. Religious background includes three dummy variables for a student's religious affiliation (Catholic, Protestant, other religion or not religious). Family background includes dummy variables for student's SES as listed in Table 1. Standard errors are reported in parentheses and are clustered at the classroom level. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

**Table 4.** In-Group Bias: How Type of Diversity Matters

| Dependent variable: | in-group out-group investment gap | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Muslim immigrant share ($\pi_{D,k}$) | 4.261*** | 4.168*** | 3.973*** |
| | (1.047) | (1.036) | (1.035) |
| Non-Muslim immigrant share ($\pi_{C,k}$) | 1.481** | 1.484** | 1.418** |
| | (0.699) | (0.704) | (0.711) |
| $\pi^2_{D,k}$ | -5.546*** | -5.243*** | -4.966*** |
| | (1.360) | (1.352) | (1.363) |
| $\pi^2_{C,k}$ | -0.521 | -0.697 | -0.659 |
| | (1.230) | (1.254) | (1.282) |
| $\pi_{D,k} \times \pi_{C,k}$ | -14.888** | -14.127** | -12.948** |
| | (6.381) | (6.329) | (6.373) |
| $\pi^2_{D,k} \times \pi_{C,k}$ | 13.589** | 10.921* | 9.508 |
| | (6.526) | (6.590) | (6.652) |
| $\pi_{D,k} \times \pi^2_{C,k}$ | 8.071 | 8.824 | 8.134 |
| | (7.892) | (7.799) | (7.877) |
| p-value: coeffs. jointly equal to zero | 0.0001 | 0.0001 | 0.0001 |
| Observations | 2,168 | 2,168 | 2,168 |
| R-squared | 0.056 | 0.062 | 0.067 |
| Basic controls | ✓ | ✓ | ✓ |
| Religious background | | ✓ | ✓ |
| Family background | | | ✓ |

OLS estimates of equation (5). The dependent variable is the in-group/oug-group investment gap, normalized to be mean 0 and standard deviation 1. Basic controls include school fixed effects, a student's gender and age, and class size. Religious background includes three dummy variables for a student's religious affiliation (Catholic, Protestant, other religion or not religious). Family background includes dummy variables for a student's SES as listed in Table 1. Standard errors are reported in parentheses and are clustered at the classroom level. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

**Table 5.** In-Group Bias: Classroom Polarization versus Classroom Fractionalization

| Dependent variable: | in-group out-group investment gap | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Classroom polarization | 0.605*** | | 0.539*** |
| | (0.145) | | (0.193) |
| Classroom fractionalization | | 0.402*** | 0.103 |
| | | (0.139) | (0.181) |
| Observations | 2,168 | 2,168 | 2,168 |
| R-squared | 0.064 | 0.062 | 0.062 |
| Basic controls | ✓ | ✓ | ✓ |
| Religious background | ✓ | ✓ | ✓ |
| Family background | ✓ | ✓ | ✓ |

OLS estimates. The dependent variable is the in-group/oug-group investment gap, normalized to be mean 0 and standard deviation 1. Classroom polarization and fractionalization are defined using equation 2 and the 11 origin countries given in Table A1. All regressions include the full set of control variables used in column (3) of Table 3. Standard errors are reported in parentheses and are clustered at the classroom level. ***,**,* indicate significance at the 1%, 5%, and 10% level, respectively.

# Appendix for Online Publication

"Diversity and Cooperation"

by Dan Anderberg, Gordon B. Dahl, Christina Felfe, Helmut Rainer, and Thomas Siedler

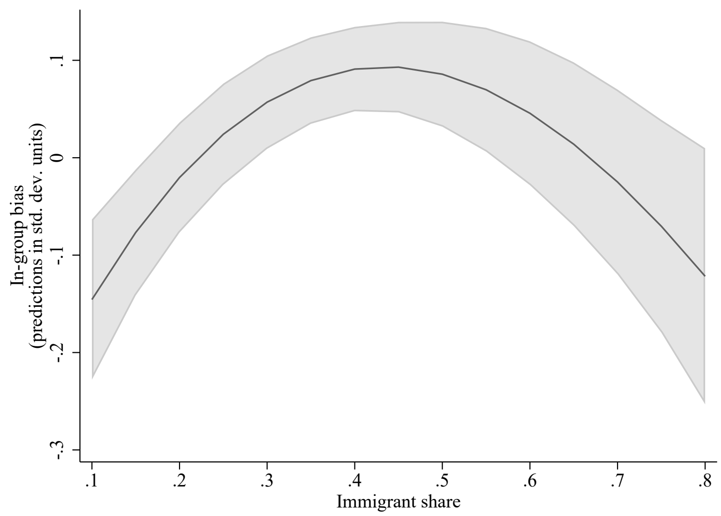
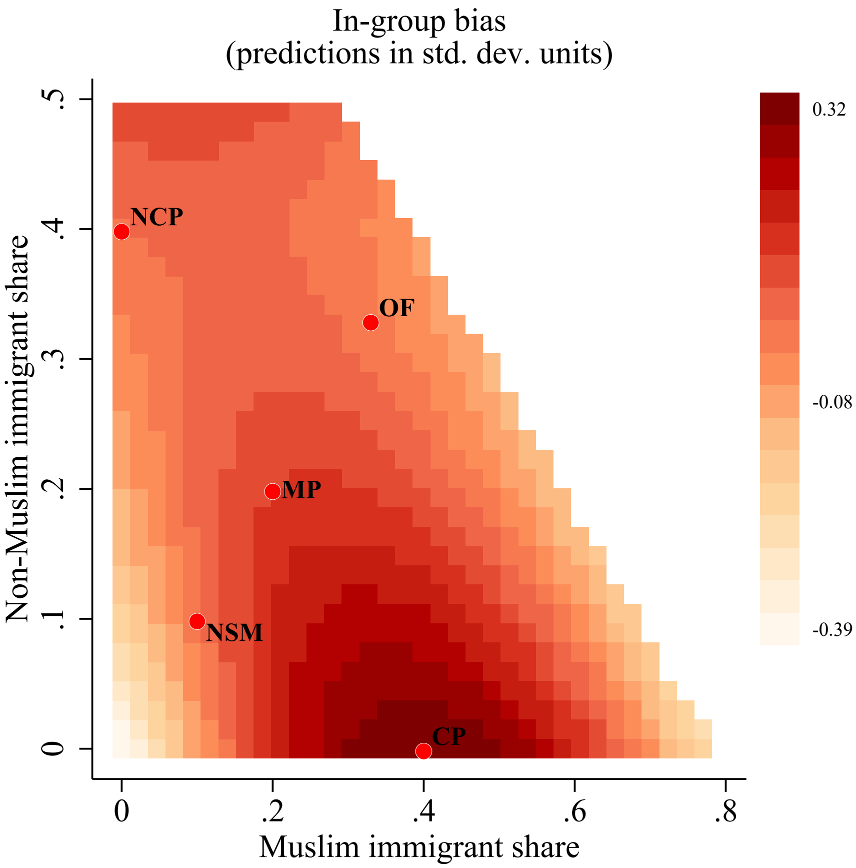



**Figure A1.** In-Group Bias Using a Cubic Specification

*Notes:* The figure shows predictions similar to those in Figure 4, but with diversity modeled as a cubic polynomial in the the proportion of immigrant peers. The grey shaded area denotes pointwise 90% confidence intervals. N=2,168. 1 standard deviation = 76 euro-cents.




| Differences between classroom types (in std. dev. units) | |
|---|---|
| CP-NCP | 0.321** |
| | (0.156) |
| CP-MP | 0.224 |
| | (0.168) |
| CP-OF | 0.339* |
| | (0.177) |
| CP-NSM | 0.329** |
| | (0.157) |

**Figure A2.** Robustness: Third-Order Expansion Estimates for In-Group Bias

*Notes:* In-group bias is measured by the in-group out-group investment gap in the first stage of the investment game for natives. The heatmap shows predicted values for in-group bias using a complete third-order expansion of the non-Muslim and Muslim immigrant fractions in a classroom. The peak in in-group bias occurs in a classroom exhibiting **cultural polarization (CP)**, with native German peers forming a slight majority group [60%] and Muslim immigrant peers a large minority group [40%]. The table shows the difference between predicted in-group bias at this peak and four other classroom scenarios:

**Non-cultural polarization (NCP)**: [60% natives, 40% non-Muslim immigrants];

**Mixed polarization (MP)**: [60% natives, 20% Muslim immigrants, 20% non-Muslim immigrants];

**Overall fractionalization (OF)**: [34% natives, 33% Muslim immigrants, 33% non-Muslim immigrants];

**Native supermajority (NSM)**:[80% natives, 10% Muslim immigrants, 10% non-Muslim immigrants].

N=2,168. 1 standard deviation = 76 euro-cents. *** p<0.01, ** p<0.05, * p<0.1.

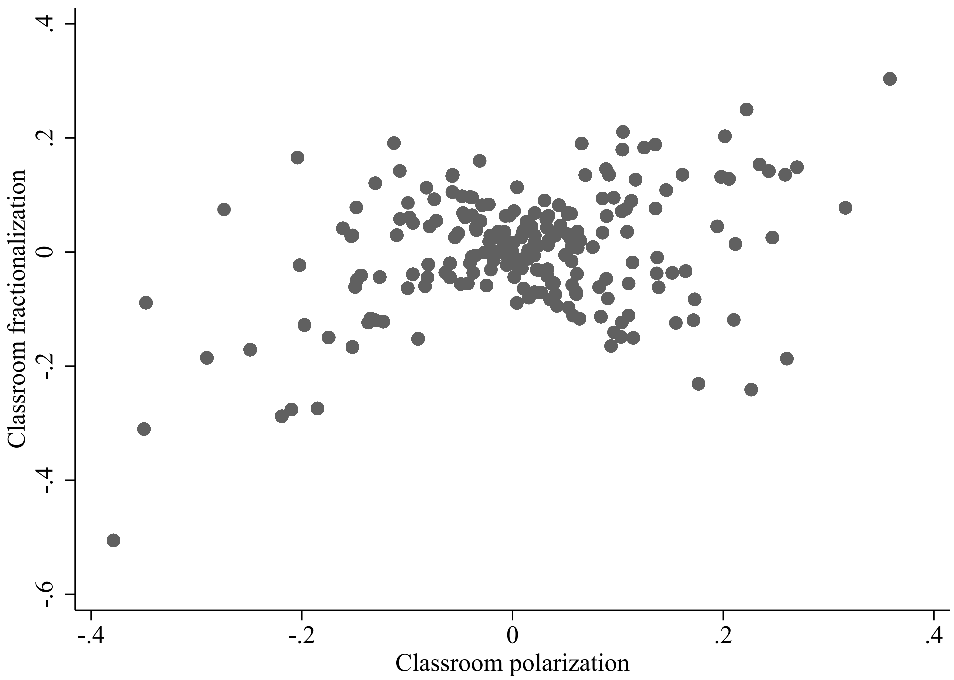



**Figure A3.** Polarization and Fractionalization: Identifying Variation after Removing School Fixed Effects

*Notes:* Classroom polarization and fractionalization are defined using equation 2 and the 11 origin countries given in Table A1. The figure shows a scatter plot of the residualized fractionalization index against the residualized polarization index, both net of school fixed effects.

**Table A1.** Origin Countries of Peers, Mapped into 11 Regions

| | Mean | St. Dev. |
|---|---|---|
| Immigrant share | 0.391 | 0.217 |
| *Fraction of immigrant peers from:* | | |
| Turkey | 0.134 | 0.139 |
| Balcan countries | 0.036 | 0.048 |
| Eastern European countries | 0.044 | 0.052 |
| Post Soviet countries | 0.039 | 0.049 |
| Southern European countries | 0.019 | 0.039 |
| Central and Northern European countries | 0.013 | 0.030 |
| Midlle Eastern countries | 0.024 | 0.041 |
| Asian countries | 0.022 | 0.036 |
| African countries | 0.023 | 0.043 |
| Rest of the world | 0.010 | 0.025 |
| Unidentified countries | 0.017 | 0.034 |
| Observations | 2,168 | |

*Notes*: Origin country of peers defined by parent's country of origin. See text for details.

# Appendix B: A Framework

In this appendix we present further details of the framework described in Section 6.

## 1.1 Setup and Equilibrium Group Joining Rates

The setup is as outlined in Section 6 and is repeated here only for completeness. Consider an economy with a continuum of individuals who are of $J \geq 2$ types. Let the proportions of types in the population be denoted $\boldsymbol{\pi} = (\pi_1, ..., \pi_J)$. $J+1$ groups form endogenously. For each type $j$ there is an *exclusive* group comprising only $j$-type individuals. The final group is an *inclusive* group with members of all types. Individual $i$ of type $j$, chooses between two options: join the own-type exclusive group (option 0) or join the inclusive group (option 1). Joining the inclusive group involves a type-specific joining cost $h_j$ which below will be related to cultural distance. Finally, let $\beta > 0$ parameterize the (common) strength of preference for group size. The individual's utilities of the two options are,

$$u_{ij}^0 = \beta \pi_j (1 - \mu_j) + \varepsilon_{ij}^0, \qquad u_{ij}^1 = \beta \sum_{j'=1}^{J} \pi_{j'} \mu_{j'} - h_j + \varepsilon_{ij}^1, \tag{B.1}$$

where $\mu_j$ is the proportion of type-$j$ individuals who join the inclusive group, and where we used that the size of the type-$j$ exclusive group is $\pi_j (1 - \mu_j)$ while the size of the inclusive group is $\sum_{j'=1}^{J} \pi_{j'} \mu_{j'}$. $\varepsilon_{ij}^0$ and $\varepsilon_{ij}^1$ are the individual's choice-specific random preferences. To rule out multiple coordination equilibria we place an upper limit on $\beta$, and we also make a specific, but standard, distributional assumption on the random preferences.

**Assumption 1.** *(Preferences) The preference for group-size satisfies $\beta \in (0, 2)$. Any individual i's random preferences $\varepsilon_{ji}^0$ and $\varepsilon_{ji}^1$ are i.i.d. extreme value distributed.*

With the individual random preferences being extreme value distributed, the proportion $\mu_j$ of type $j$ individuals who join the inclusive group satisfies

$$\rho_j(\boldsymbol{\mu}) \equiv \log\left(\frac{\mu_j}{1 - \mu_j}\right) - \beta \left(\sum_{j'=1}^{J} \pi_{j'} \mu_{j'} - \pi_j (1 - \mu_j)\right) + h_j = 0, \qquad j = 1, ..., J, \tag{B.2}$$

where $\boldsymbol{\mu} = (\mu_1, ..., \mu_J)$. Similarly, let $\boldsymbol{h} = (h_1, ..., h_J)$ which, for the moment, we take as given.

The inclusive group joining rate of any one type affects the joining incentives of all other types. Hence an equilibrium in joining rates requires both "within-type" consistency (for each type $j$, $\mu_j$ should satisfy (B.2) for that type given the joining rates of all other types) and "across-type" consistency (within-type consistency should hold for all $J$ types simultaneously).

**Definition 1.** *A "joining equilibrium", given the type distribution $\boldsymbol{\pi}$ and the type-specific joining costs $\boldsymbol{h}$, is a vector $\boldsymbol{\mu}$ that satisfies (B.2) for all $J$ types simultaneously.*

The following notes that a unique joining equilibrium exists.

**Proposition 1.** *(Existence of joining equilibrium). For any given type distribution $\boldsymbol{\pi}$ and type-specific joining costs $\boldsymbol{h}$, a unique joining equilibrium $\boldsymbol{\mu}$ exists.*

**Proof**. We start by showing existence. For any type $j$ and any given $\boldsymbol{\mu}_{-j}$ (the vector $\boldsymbol{\mu}$ without the $j$'th component), the equation $\rho_j(\boldsymbol{\mu}) = 0$ has a unique solution $\mu_j \in (0, 1)$. To see this, note that, for any $\boldsymbol{\mu}_{-j}$, $\rho_j(\boldsymbol{\mu})$ is continuously differentiable in $\mu_j$ and goes to $-\infty$ as $\mu_j \to 0$ and to $+\infty$ as $\mu_j \to 1$. Moreover, differentiating (B.2) gives that,

$$\frac{\partial \rho_j}{\partial \mu_j} = S_j \equiv \frac{1}{\mu_j(1-\mu_j)} - 2\beta\pi_j > 0, \qquad j = 1, .., J, \tag{B.3}$$

where the sign follows from the fact that $\mu_j(1-\mu_j) \leq 1/4$ and $\beta \in (0,2)$ (Assumption 1). This establishes that a unique solution $\mu_j \in (0,1)$ to $\rho_j(\boldsymbol{\mu}) = 0$ given $\boldsymbol{\mu}_{-j}$ exists. Denote this unique solution by $\widehat{\mu}_j\left(\boldsymbol{\mu}_{-j}\right)$. By the implicit function theorem $\widehat{\mu}_j(\cdot)$ is also a continuous function. Extending the argument of $\widehat{\mu}_j(\cdot)$ to $\boldsymbol{\mu}$ (where it is understood that $\widehat{\mu}_j(\cdot)$ only depends on $\boldsymbol{\mu}_{-j}$, not on $\mu_j$) and forming $\widehat{\boldsymbol{\mu}}(\boldsymbol{\mu}) = (\widehat{\mu}_1(\boldsymbol{\mu}), ..., \widehat{\mu}_J(\boldsymbol{\mu}))$ we thus have a continuous function that maps $[0,1]^J$ into itself. A fixed point of this mapping is a joining equilibrium and by Brouwer's theorem, such a fixed point is guaranteed to exist.

Consider next uniqueness. Let $\boldsymbol{\rho}(\boldsymbol{\mu}) = (\rho_1(\boldsymbol{\mu}), ..., \rho_J(\boldsymbol{\mu}))$ and note that a joining equilibrium is a $\boldsymbol{\mu}$ such that $\boldsymbol{\rho}(\boldsymbol{\mu}) = \mathbf{0}$, the null vector of length $J$. If the mapping $\boldsymbol{\rho}(\boldsymbol{\mu})$ can be shown to be univalent, then uniqueness follows since the null vector can then have at most one pre-image (and given existence we know that it has at least one pre-image). Gale and Nikaido (1965) showed that if the Jacobian matrix of a map $\boldsymbol{\rho}(\cdot)$ is a P-matrix, then $\boldsymbol{\rho}(\cdot)$ is univalent. A matrix $\boldsymbol{A} \in \mathbb{R}^{J \times J}$ is a P-matrix if all its principal minors are positive, and a sufficient condition for this is the dominant diagonal condition: $|a_{jj}| \geq \sum_{j' \neq j} |a_{jj'}|$ for all $j$. That is, for each row in $\boldsymbol{A}$, the absolute value of the diagonal term $a_{jj}$ is no less than the sum of the absolute values of the off-diagonal terms in that row. Using that the diagonal terms of the Jacobian of $\boldsymbol{\rho}(\boldsymbol{\mu})$ take the form (B.3) and the off-diagonal terms are $\partial \rho_j / \partial \mu_{j'} = -\beta \pi_{j'}$, $j' \neq j$, the Jacobian of $\boldsymbol{\rho}(\cdot)$ has a dominant diagonal if, for all $j$, $S_j \geq \beta \sum_{j' \neq j} \pi_{j'}$. But, since $\sum_{j' \neq j} \pi_{j'} = 1 - \pi_j$, this is equivalent to $[\mu_j(1-\mu_j)]^{-1} \geq \beta(1+\pi_j)$ which holds since the left hand side is not less than four and the right hand side is, using Assumption 1, strictly less than four. #

## 1.2 Cultural Identities and Inclusive-Group Joining Costs

There is a space of cultural identities which we take to be $\Theta \equiv [-1,+1]^{J-1}$. Each type $j$ is endowed with some *exogenously* given identity $\boldsymbol{\theta}_j \in \Theta$. The inclusive group has an *endogenous* identity $\boldsymbol{\theta} \in \Theta$, making the type-specific (Euclidean) distance $d_j \equiv \|\boldsymbol{\theta} - \boldsymbol{\theta}_j\|$. We assume a type-specific cost of joining the inclusive group that depends on this cultural distance: $h_j = h_j(d_j)$ where, for each type $j$, $h_j(\cdot)$ is twice continuously differentiable and satisfies $h_j(0) = 0$, $h_j'(\cdot) > 0$ and $h_j''(\cdot) > 0$. To close the model we need to specify how $\boldsymbol{\theta}$ is determined. In line with the assumption of positive preferences for group size, we assume that $\boldsymbol{\theta}$ is chosen so as to maximize the inclusive group's size (or "popularity") in the joining equilibrium that ensues.

**Definition 2.** *(Inclusive group identity) The cultural identity of the inclusive group* $\boldsymbol{\theta} \in \Theta$ maximizes *the size of the inclusive group,* $\boldsymbol{\theta} = \arg\max_{\boldsymbol{\theta} \in \Theta} \Omega(\theta)$*, where* $\Omega(\theta) \equiv \sum_{j=1}^{J} \pi_j \mu_j$.

If the types have different population frequencies, then it is natural that $\boldsymbol{\theta}$ is set close to the endowed identity of a relatively more frequent type. However, the type specific aversions to distance also matter. The case of just two types allows us to illustrate this.

### 1.2.1 Comparative Statics with Two Types

Consider the $J = 2$ case: natives (type 1) and immigrants (type 2). With $J = 2$, the identity space is $\Theta = [-1,+1]$ and we can assume that the endowed cultural identities of natives and immigrants are $\theta_1 = -1$ and $\theta_2 = +1$ respectively. The identity of the inclusive group is a scalar $\theta \in \Theta$ whereby $d_1 = 1 + \theta$ and $d_2 = 1 - \theta$, with corresponding type-specific joining costs $h_1 = h_1(1+\theta)$ and $h_2 = h_2(1-\theta)$. The inclusive group identity $\theta$ maximizes $\Omega(\theta) \equiv \sum_{j=1}^{2} \pi_j \mu_j$ where the joining rates $\boldsymbol{\mu} = (\mu_1, \mu_2)$ simultaneously satisfy (B.2) for $j = 1, 2$.

We will consider deviations from "symmetry" where the two types are of equal proportion in the population, $\pi_1 = \pi_2 = 1/2$, and have the same distance cost function, $h_j(\cdot) = h(\cdot)$ for some common $h(\cdot)$. Symmetric population proportions and costs also imply that $\theta = 0$ is a stationary point of $\Omega(\theta)$ (see below) at which both types have the same inclusive group joining rate, $\mu^*$, satisfying

$$\log\left(\frac{\mu^*}{1-\mu^*}\right) - \frac{\beta}{2}(3\mu^* - 1) + h(1) = 0. \tag{B.4}$$

Consider first how $\mu_1$ and $\mu_2$ are affected by $\theta$ at symmetry? Simple comparative statics show that

$$\left.\frac{\partial \mu_2}{\partial \theta}\right|_{sym} = -\left.\frac{\partial \mu_1}{\partial \theta}\right|_{sym} = \frac{h'(1)}{S^* + C^*} > 0, \tag{B.5}$$

where $S^* = [\mu^*(1-\mu^*)]^{-1} - \beta > 0$ and $C^* = \beta/2 > 0$, whereby $S^* + C^* = [\mu^*(1-\mu^*)]^{-1} - \beta/2 > 0$. This trivially confirms that $\theta = 0$ is indeed a stationary point of $\Omega(\theta)$ under symmetric population proportions and distance costs[23]

$$\left.\frac{\partial \Omega(\theta)}{\partial \theta}\right|_{sym} = \frac{1}{2}\sum_{j=1}^{2} \left.\frac{\partial \mu_j}{\partial \theta}\right|_{sym} = 0. \tag{B.6}$$

Similarly, at symmetry, and using $\pi_1 = 1 - \pi_2$,

$$\left.\frac{\partial \mu_1}{\partial \pi_2}\right|_{sym} = -\left.\frac{\partial \mu_2}{\partial \pi_2}\right|_{sym} = \frac{\beta(1-\mu^*)}{S^* + C^*} > 0. \tag{B.7}$$

Turning to the cross-partials, first for the equilibrium joining rates, it can be shown that

$$\left.\frac{\partial^2 \mu_j}{\partial \theta \partial \pi_2}\right|_{sym} = \frac{h'(1)}{D^*}\left(\frac{(2\mu^* - 1)}{(\mu^*)^2(1-\mu^*)^2}\frac{\beta(1-\mu^*)}{(S^* + C^*)} + 3\beta\right), \tag{B.8}$$

where $D^* = (S^* + C^*)(S^* - C^*) > 0$. More importantly for our purposes however, the cross partial of $\Omega(\theta)$ at symmetry (after substituting using (B.5), and (B.8) and also using the expressions for $S^*$, $C^*$ and $D^*$) can be shown to be

$$\left.\frac{\partial \Omega(\theta)}{\partial \theta \partial \pi_2}\right|_{sym} = \frac{8\mu^*(1-\mu^*)((2-\beta) + \beta\mu^*(2-\mu^*))h'(1)}{(2 - \beta\mu^*(1-\mu^*))^2(2 - 3\beta\mu^*(1-\mu^*))} > 0, \tag{B.9}$$

where the sign follows from $\beta \in (0, 2)$ (Assumption 1). As a result, it follows that $\theta$ is increasing in $\pi_2$ at symmetry. That is, when type 2 (type 1) becomes a majority (minority) type, the inclusive group's identity moves closer to type 2 (away from type 1).

[23] Different sufficient conditions can be given for $\theta = 0$ to be a (local) maximum of $\Omega(\theta)$ under symmetric type proportions and distance costs. Either sufficient convexity of the common cost function, $h''(1)/(h'(1))^2 > 1$ (for any $\mu^*$) or any convexity $h''(1) > 0$ and $\mu^* > 1/2$ (where the latter would be implied by $h(1) < \beta/4$).

Consider next the effect of introducing type-differences in the distance cost. Specifically, let $h_1(d) = (1-\epsilon) h(d)$ and $h_2(d) = (1+\epsilon) h(d)$, where we then consider the introduction of $\epsilon > 0$ from symmetry ($\epsilon = 0$). It can then be shown that,

$$\left.\frac{\partial^2(\theta)}{\partial\theta\partial\epsilon}\right|_{sym} = \frac{h'(1)}{D^*}\left[(S^*+C^*) + \frac{(2\mu^*-1)}{(\mu^*)^2(1-\mu^*)^2}\frac{h(1)}{(S^*+C^*)}\right]. \tag{B.10}$$

The main effect here – captured by the first component – is that $\epsilon > 0$ increases the cost for type 2 relative to type 1, whereby $\theta$ is moved closer to type 2 to retain participation from this group. The second component in (B.10) (which is also positive whenever $\mu^* > 1/2$) reflects an impact on the relative size of the social multipliers, $S_1$ and $S_2$.

### 1.2.2 Illustrative Example: Limiting Cases

The example provided in Section 6 had three types: natives ($N$), culturally close immigrants ($C$), culturally distant immigrants ($D$). The overall immigrant share was denoted by $\pi$ and the split between the immigrant types was denoted $s$, whereby $\pi_N = 1-\pi$, $\pi_C = \pi(1-s)$ and $\pi_D = \pi s$. When either $s \to 0$ or $s \to 1$ the example limits to just natives and one immigrant type. In the former limiting case, the only immigrant type to exist is the culturally close type. In the latter limiting case, the only immigrant type to exist is the culturally distant type. Here we showcase how the example given also provides a good correspondence to our empirical findings in these limiting cases. For instance, increasing the immigrant share $\pi$ with $s = 0$, in equivalent to moving on a vertical ray out from the origin in our heatmaps, that is, along the $y$-axis. Similarly, increasing $\pi$ whilst for $s = 1$, in equivalent to moving on a horizontal ray out from the origin in our heatmaps, that is, along the $x$-axis.

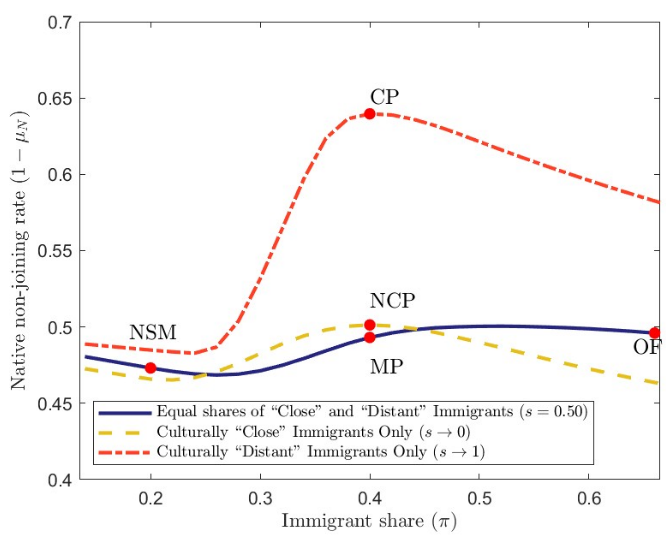



**Figure B.1.** Native Inclusive Group Non-Joining Rate, $1-\mu_N$, in the Limiting Cases

*Notes:* Native inclusive group non-joining rate, $1-\mu_N$, in the case of equal proportions of culturally close and culturally distant immigrant types $s = 1/2$, and in the limiting cases with only culturally close immigrants, $s \to 0$, and only culturally distant immigrants, $s \to 1$, respectively. Model parameters set at $\gamma_N = 0.3$, $\gamma_C = \gamma_D = 0.8$, $\beta = 1.1$ and $\sigma = 1.2$.

Figure **??** illustrates the equilibrium native inclusive group non-joining rate, $1-\mu_n$, in these two limiting cases for the particular parameterization used in Section 6. The blue solid line is for $s = 1/2$ as already shown in Figure 13. The yellow dashed line is for the limiting case where $s \to 0$ and thus shows how the native inclusive group non-joining rate varies with the immigrant share when the only immigrant type is the culturally close type. In line with the empirical findings, the native non-joining rate increases when $\pi$ increases from zero, giving rise to non-cultural polarization ($NCP$, corresponding to $\pi = 0.4$ with $s = 0$). However, the increase in the native non-joining rate in this case is modest relative to that which occurs under cultural polarization ($CP$, corresponding to $\pi = 0.4$ with $s = 1$). Indeed, in the limiting case with only culturally distant immigrants, $s \to 1$, the native non-joining rate exhibits a strong inverted U-shape as the immigrant share $\pi$ is varied, directly matching the inverted U-shape in in-group bias when moving along the along the $x$-axis in Figure 4.

## Appendix C: Translation of Experimental Instructions and Decision Sheets

*Experimental Instructions*

Participation in this game is voluntary!

Thank you very much for participating. From now on, please do not speak with anyone else apart from us about the game. Unfortunately, if you break this rule, we will have to exclude you from the game.

The objective of this game is to examine how people make decisions. There are no “right” or “wrong” decisions in the game and our aim is not to test your knowledge. Make your decisions exactly as you wish. During this game, you will be earning real money. We guarantee that you will receive a cash payout within two weeks. You will receive your money in an envelope marked with your ID number, so please make sure you keep your ID number in a safe place! These envelopes will be passed out by one of your teachers or can be collected from the secretary’s office.

The amount of money you earn depends on your decisions and the decisions of the other participants. We will now describe the rules in detail. It is therefore especially important that you listen very carefully.

There are no “right” or “wrong” decisions in this game. You should make your decisions based on your own personal deliberations. Your decisions will remain anonymous, which means that no one else will know what you decide.

If you have any questions after reading these instructions, please raise your hand. Someone will then come over to you and answer your questions in private (i.e., quietly).

**Process:**

There are two roles in this game: **sender** and **responder**.

The game starts as follows: Each sender and each responder receives 5 euros. The sender must decide how much of the 5 euros he/she wishes to give to the responder.

The amount the sender gives to “his/her” responder will then be tripled. In other words, the responder receives precisely three times the amount the sender has given him/her.

Next, it is the responder’s turn. He/she now has three times the amount the sender has given him/her plus his/her own 5 euros. The responder must now decide how much of this money he/she would like to return to “his/her” sender. Please note: The sum the responder returns to the sender is not tripled.

**Payment:**

At the end of the game, the sender receives the sum that he/she kept plus the sum that the responder returned to him/her.

**Payment to sender = 5 euros - sum sent + sum returned (by responder)**

The responder receives the sum he/she was given by the sender (times 3), minus the sum he/she returned to the sender.

**Payment to responder = 5 euros + 3 x sum sent (by sender) - sum returned**

**Decisions:**

You will be required to make one decision in the role of sender and one in the role of responder. You can also choose between different "categories" of senders and responders; you obviously do not have to treat these groups differently, however. These categories are described on the decision sheet. You can, for instance, choose whether you send or return money to a boy or a girl. It is your decision, there is no "right" or "wrong".

**Calculating your payment:**

Some of the following points will be easier to understand once you have seen the decision sheets. We will now go through the points and then look at the decision sheets together. If you still have questions after that, we will be happy come back to these points.

Once the game has been carried out in several schools, the following will happen:

1. Two students from different schools will be randomly paired; you will therefore not know "your" sender or "your" responder personally; however, he or she will be around the same age as you and will also go to school in North Rhine-Westphalia.

2. Who is to play the role of the sender and who the role of the responder will also be randomly decided.

3. Next, we identify the category (see decision sheet) that the sender and responder are each from. This information is extracted from the questionnaire you completed. The sender can be a girl and the responder a boy, for instance.

4. Next, the sender's decision is implemented based on the actual category of the responder.

5. Finally, the responder's decision is implemented based on the actual category of the sender and the actual amount received from "their" sender.

6. We now know how much the sender has sent and how much the responder has returned. Based on this, we can calculate the payment to both the sender and the responder. This money is then placed in the appropriate envelopes marked with the corresponding sender and responder ID numbers and taken to the schools.

7. At the end, you will be able to collect the envelope containing your payment at your school.

Now look at the decision sheets. This will help you to better understand some of the points described above. Think carefully about the decisions you wish to make. You have plenty of time! If you have any questions, please raise your hand. Someone will then come over to you and answer your questions in private (i.e., quietly).

**ID: □□□□□□**

✂ - - - - - - - - - - - - - - - - - - - - - ✂

**Please KEEP your ID!!!!!**

**ID: □□□□□□**

# Sender

**You are the sender and you have 5 EURO. Which amount would you like to send to the receiver (max. 5 EURO)? Please check one box in each column 1-6.**

| **The receiver is...** | | | | | |
|---|---|---|---|---|---|
| **COLUMN 1** | **COLUMN 2** | **COLUMN 3** | **COLUMN 4** | **COLUMN 5** | **COLUMN 6** |
| *... a boy with German parents* | *... a girl with German parents* | *... a boy with foreign parents* | *... a girl with foreign parents* | *... a boy with foreign parents who possesses German citizenship* | *... a girl with foreign parents who possesses German citizenship* |
| ☐ 0 EURO | ☐ 0 EURO | ☐ 0 EURO | ☐ 0 EURO | ☐ 0 EURO | ☐ 0 EURO |
| ☐ 0.5 EURO | ☐ 0.5 EURO | ☐ 0.5 EURO | ☐ 0.5 EURO | ☐ 0.5 EURO | ☐ 0.5 EURO |
| ☐ 1 EURO | ☐ 1 EURO | ☐ 1 EURO | ☐ 1 EURO | ☐ 1 EURO | ☐ 1 EURO |
| ☐ 1.5 EURO | ☐ 1.5 EURO | ☐ 1.5 EURO | ☐ 1.5 EURO | ☐ 1.5 EURO | ☐ 1.5 EURO |
| ☐ 2 EURO | ☐ 2 EURO | ☐ 2 EURO | ☐ 2 EURO | ☐ 2 EURO | ☐ 2 EURO |
| ☐ 2.5 EURO | ☐ 2.5 EURO | ☐ 2.5 EURO | ☐ 2.5 EURO | ☐ 2.5 EURO | ☐ 2.5 EURO |
| ☐ 3 EURO | ☐ 3 EURO | ☐ 3 EURO | ☐ 3 EURO | ☐ 3 EURO | ☐ 3 EURO |
| ☐ 3.5 EURO | ☐ 3.5 EURO | ☐ 3.5 EURO | ☐ 3.5 EURO | ☐ 3.5 EURO | ☐ 3.5 EURO |
| ☐ 4 EURO | ☐ 4 EURO | ☐ 4 EURO | ☐ 4 EURO | ☐ 4 EURO | ☐ 4 EURO |
| ☐ 4.5 EURO | ☐ 4.5 EURO | ☐ 4.5 EURO | ☐ 4.5 EURO | ☐ 4.5 EURO | ☐ 4.5 EURO |
| ☐ 5 EURO | ☐ 5 EURO | ☐ 5 EURO | ☐ 5 EURO | ☐ 5 EURO | ☐ 5 EURO |

**How much do you think "your" receiver will send back to you? Consider that he or she now has 5 EURO plus three times the amount that you sent (max. 20 EURO). Recall: We tripled the amount you sent. Please fill in columns 1-6 and use one decimal place at most (10-cent steps).**

| **The receiver is...** | | | | | |
|---|---|---|---|---|---|
| **COLUMN 1** | **COLUMN 2** | **COLUMN 3** | **COLUMN 4** | **COLUMN 5** | **COLUMN 6** |
| *... a boy with German parents* | *... a girl with German parents* | *... a boy with foreign parents* | *... a girl with foreign parents* | *... a boy with foreign parents who possesses German citizenship* | *... a girl with foreign parents who possesses German citizenship* |
| _____ EURO | _____ EURO | _____ EURO | _____ EURO | _____ EURO | _____ EURO |

# Receiver 1

**You are the receiver. The sender is a <u>boy with German parents.</u> How much do you want to send back to him? Please fill in an amount for each possible case (at most one decimal place = 10-cent steps.)**

| *Assume the sender has sent you the following amount:* | *The sender still has:* | *You have:* | *Which amount do you want to send back:* | *Potential amount to send back:* |
|---|---|---|---|---|
| 0 EURO | 5 EURO | 5 EURO | _____ EURO | (0 to 5 EURO) |
| 0.5 EURO | 4.5 EURO | 6.5 EURO | _____ EURO | (0 to 6.5 EURO) |
| 1 EURO | 4 EURO | 8 EURO | _____ EURO | (0 to 8 EURO) |
| 1.5 EURO | 3.5 EURO | 9.5 EURO | _____ EURO | (0 to 9.5 EURO) |
| 2 EURO | 3 EURO | 11 EURO | _____ EURO | (0 to 11 EURO) |
| 2.5 EURO | 2.5 EURO | 12.5 EURO | _____ EURO | (0 to 12.5 EURO) |
| 3 EURO | 2 EURO | 14 EURO | _____ EURO | (0 to 14 EURO) |
| 3.5 EURO | 1.5 EURO | 15.5 EURO | _____ EURO | (0 to 15.5 EURO) |
| 4 EURO | 1 EURO | 17 EURO | _____ EURO | (0 to 17 EURO) |
| 4.5 EURO | 0.5 EURO | 18.5 EURO | _____ EURO | (0 to 18.5 EURO) |
| 5 EURO | 0 EURO | 20 EURO | _____ EURO | (0 to 20 EURO) |

# Receiver 2

**You are the receiver. The sender is a girl with German parents. How much do you want to send back to her? Please fill in an amount for each possible case (at most one decimal place = 10-cent steps.)**

| *Assume the sender has sent you the following amount:* | *The sender still has:* | *You have:* | *Which amount do you want to send back:* | *Potential amount to send back:* |
|---|---|---|---|---|
| 0 EURO | 5 EURO | 5 EURO | _____ EURO | (0 to 5 EURO) |
| 0.5 EURO | 4.5 EURO | 6.5 EURO | _____ EURO | (0 to 6.5 EURO) |
| 1 EURO | 4 EURO | 8 EURO | _____ EURO | (0 to 8 EURO) |
| 1.5 EURO | 3.5 EURO | 9.5 EURO | _____ EURO | (0 to 9.5 EURO) |
| 2 EURO | 3 EURO | 11 EURO | _____ EURO | (0 to 11 EURO) |
| 2.5 EURO | 2.5 EURO | 12.5 EURO | _____ EURO | (0 to 12.5 EURO) |
| 3 EURO | 2 EURO | 14 EURO | _____ EURO | (0 to 14 EURO) |
| 3.5 EURO | 1.5 EURO | 15.5 EURO | _____ EURO | (0 to 15.5 EURO) |
| 4 EURO | 1 EURO | 17 EURO | _____ EURO | (0 to 17 EURO) |
| 4.5 EURO | 0.5 EURO | 18.5 EURO | _____ EURO | (0 to 18.5 EURO) |
| 5 EURO | 0 EURO | 20 EURO | _____ EURO | (0 to 20 EURO) |

# Receiver 3

**You are the receiver. The sender is a <u>boy with foreign parents.</u> How much do you want to send back to him? Please fill in an amount for each possible case (at most one decimal place = 10-cent steps.)**

| *Assume the sender has sent you the following amount:* | *The sender still has:* | *You have:* | *Which amount do you want to send back:* | *Potential amount to send back:* |
|---|---|---|---|---|
| 0 EURO | 5 EURO | 5 EURO | _____ EURO | (0 to 5 EURO) |
| 0.5 EURO | 4.5 EURO | 6.5 EURO | _____ EURO | (0 to 6.5 EURO) |
| 1 EURO | 4 EURO | 8 EURO | _____ EURO | (0 to 8 EURO) |
| 1.5 EURO | 3.5 EURO | 9.5 EURO | _____ EURO | (0 to 9.5 EURO) |
| 2 EURO | 3 EURO | 11 EURO | _____ EURO | (0 to 11 EURO) |
| 2.5 EURO | 2.5 EURO | 12.5 EURO | _____ EURO | (0 to 12.5 EURO) |
| 3 EURO | 2 EURO | 14 EURO | _____ EURO | (0 to 14 EURO) |
| 3.5 EURO | 1.5 EURO | 15.5 EURO | _____ EURO | (0 to 15.5 EURO) |
| 4 EURO | 1 EURO | 17 EURO | _____ EURO | (0 to 17 EURO) |
| 4.5 EURO | 0.5 EURO | 18.5 EURO | _____ EURO | (0 to 18.5 EURO) |
| 5 EURO | 0 EURO | 20 EURO | _____ EURO | (0 to 20 EURO) |

# Receiver 4

**You are the receiver. The sender is a girl with foreign parents. How much do you want to send back to her? Please fill in an amount for each possible case (at most one decimal place = 10-cent steps.)**

| *Assume the sender has sent you the following amount:* | *The sender still has:* | *You have:* | *Which amount do you want to send back:* | *Potential amount to send back:* |
|---|---|---|---|---|
| 0 EURO | 5 EURO | 5 EURO | _____ EURO | (0 to 5 EURO) |
| 0.5 EURO | 4.5 EURO | 6.5 EURO | _____ EURO | (0 to 6.5 EURO) |
| 1 EURO | 4 EURO | 8 EURO | _____ EURO | (0 to 8 EURO) |
| 1.5 EURO | 3.5 EURO | 9.5 EURO | _____ EURO | (0 to 9.5 EURO) |
| 2 EURO | 3 EURO | 11 EURO | _____ EURO | (0 to 11 EURO) |
| 2.5 EURO | 2.5 EURO | 12.5 EURO | _____ EURO | (0 to 12.5 EURO) |
| 3 EURO | 2 EURO | 14 EURO | _____ EURO | (0 to 14 EURO) |
| 3.5 EURO | 1.5 EURO | 15.5 EURO | _____ EURO | (0 to 15.5 EURO) |
| 4 EURO | 1 EURO | 17 EURO | _____ EURO | (0 to 17 EURO) |
| 4.5 EURO | 0.5 EURO | 18.5 EURO | _____ EURO | (0 to 18.5 EURO) |
| 5 EURO | 0 EURO | 20 EURO | _____ EURO | (0 to 20 EURO) |

# Receiver 5

**You are the receiver. The sender is a <u>boy with foreign parents who possesses German citizenship.</u> How much do you want to send back to him? Please fill in an amount for each possible case (at most one decimal place = 10-cent steps.)**

| *Assume the sender has sent you the following amount:* | *The sender still has:* | *You have:* | *Which amount do you want to send back:* | *Potential amount to send back:* |
|---|---|---|---|---|
| 0 EURO | 5 EURO | 5 EURO | _____ EURO | (0 to 5 EURO) |
| 0.5 EURO | 4.5 EURO | 6.5 EURO | _____ EURO | (0 to 6.5 EURO) |
| 1 EURO | 4 EURO | 8 EURO | _____ EURO | (0 to 8 EURO) |
| 1.5 EURO | 3.5 EURO | 9.5 EURO | _____ EURO | (0 to 9.5 EURO) |
| 2 EURO | 3 EURO | 11 EURO | _____ EURO | (0 to 11 EURO) |
| 2.5 EURO | 2.5 EURO | 12.5 EURO | _____ EURO | (0 to 12.5 EURO) |
| 3 EURO | 2 EURO | 14 EURO | _____ EURO | (0 to 14 EURO) |
| 3.5 EURO | 1.5 EURO | 15.5 EURO | _____ EURO | (0 to 15.5 EURO) |
| 4 EURO | 1 EURO | 17 EURO | _____ EURO | (0 to 17 EURO) |
| 4.5 EURO | 0.5 EURO | 18.5 EURO | _____ EURO | (0 to 18.5 EURO) |
| 5 EURO | 0 EURO | 20 EURO | _____ EURO | (0 to 20 EURO) |

# Receiver 6

**You are the receiver. The sender is a girl with foreign parents who possesses German citizenship. How much do you want to send back to her? Please fill in an amount for each possible case (at most one decimal place = 10-cent steps.)**

| *Assume the sender has sent you the following amount:* | *The sender still has:* | *You have:* | *Which amount do you want to send back:* | *Potential amount to send back:* |
|---|---|---|---|---|
| 0 EURO | 5 EURO | 5 EURO | _____ EURO | (0 to 5 EURO) |
| 0.5 EURO | 4.5 EURO | 6.5 EURO | _____ EURO | (0 to 6.5 EURO) |
| 1 EURO | 4 EURO | 8 EURO | _____ EURO | (0 to 8 EURO) |
| 1.5 EURO | 3.5 EURO | 9.5 EURO | _____ EURO | (0 to 9.5 EURO) |
| 2 EURO | 3 EURO | 11 EURO | _____ EURO | (0 to 11 EURO) |
| 2.5 EURO | 2.5 EURO | 12.5 EURO | _____ EURO | (0 to 12.5 EURO) |
| 3 EURO | 2 EURO | 14 EURO | _____ EURO | (0 to 14 EURO) |
| 3.5 EURO | 1.5 EURO | 15.5 EURO | _____ EURO | (0 to 15.5 EURO) |
| 4 EURO | 1 EURO | 17 EURO | _____ EURO | (0 to 17 EURO) |
| 4.5 EURO | 0.5 EURO | 18.5 EURO | _____ EURO | (0 to 18.5 EURO) |
| 5 EURO | 0 EURO | 20 EURO | _____ EURO | (0 to 20 EURO) |