Diversity and Performance in Entrepreneurial Teams*

Sophie Calder-Wang, *Wharton School, University of Pennsylvania*
Paul A. Gompers, *Harvard Business School, and NBER*
Kanyuan (Kevin) Huang, *UCLA*

October 25, 2021

Abstract

We study the role of diversity on the performance of entrepreneurial teams by exploiting a unique experimental setting of over 3,000 MBA students who participated in a business course to build startups. First, we quantify the strong selection based upon shared attributes when students are allowed to choose teammates. Team formation based upon shared endowed demographic characteristics such as gender, race, and ethnicity is stronger than team formation based upon acquired characteristics such as education and industry background. Second, when team memberships are randomly assigned, greater racial/ethnic diversity leads to significantly worse performance. Interestingly, the negative performance effect of diversity is partially alleviated in cohorts where teams are formed voluntarily. Finally, we find that teams with more female members performed substantially better when their faculty section leader was female. These findings suggest that policy interventions targeting greater diversity should consider match-specific qualities in forming teams to prevent the potential negative impact of diversity and aim to reduce existing biases against certain groups. Our results on vertical diversity suggest that capital allocators could also play an important role in the mentoring and advising of minority entrepreneurs.

JEL No. J1, J15, J16, L26, M12

---

* We are grateful to Kari Limmer at Harvard Business School MBA Student and Academic Services for providing us with extensive course information. Patrick Sweeney provided excellent research assistance. Support for this research was provided by the Division of Research at HBS. All errors are our own.

1. Introduction

Economics research on diversity has explored both impediments to team diversity (Becker, 1957; Morgan and Vardy, 2009), the performance effects of diversity on team performance (Mello and Ruckes, 2006), as well as the effects of diversity in the mentor-employee relationship and its impact on performance and promotion (Athey, Avery, and Zemsky, 2000). Most of this work has focused on theoretical models that derive predictions about factors that may hinder or promote workplace diversity or how shared attributes of supervisors and employees may aid groups with acquiring human capital and increasing the likelihood of promotion. By exploiting a series of quasi-random variations in a Harvard Business School entrepreneurship course, our paper fills the empirical gap in this research on the factors that shape team diversity and how horizontal (team-level) diversity and vertical (supervisor-team) diversity affects performance.

This paper contributes three major findings to the literature on diversity and its effects on performance. First, we quantify what factors are important for limiting diversity in forming entrepreneurial teams. Becker (1957) was the first to model factors that might lead to homogeneity in organizations in economics. Research has also documented the existence of homophily, the desire to associate with those similar to you, in various social networks, from the strongest social ties such as in marriage (Kalmijn 1998, Fiore and Donath 2005), close friendships (Marsden, 1987, 1988, Currarini, Jackson and Pin 2009), to professional networks (Gompers et al. 2016; Kleinbaum et al. 2013; Ruef et al. 2003; Reagans 2011; Sorenson and Stuart 2001) and acquaintances (Hampton and Wellman 2000). Most past research has focused on homophily in endowed demographic characteristics such as race/ethnicity and gender. However, homophily can also be based on acquired characteristics, such as education, occupation, and religion (Lazarsfeld and Merton 1954). Relatively few studies have examined homophily in educational and professional backgrounds due to limitations in data.

We estimate the relative magnitudes of homophily in race/ethnicity, gender, education, and work experience by quantifying which of these shared attributes are more likely to lead to team membership. Using a novel dataset of HBS MBA students' choices to co-found real micro-businesses, we find that team formation based upon shared endowed demographic characteristics is stronger than team formation based upon acquired characteristics. Individuals are 25% more likely to form groups with people of the same race/ethnicity or gender relative to randomly matching.

Selection based upon education history and work experience is weaker than endowed demographic attributes but still economically significant. School ties and shared work experience increase the probability of co-founding a micro-business by 17% and 11%, respectively. Further, we find that team selection effects of shared education and work experience are stronger among male students than female students. We are able to quantify the relative strengths of homophily forces because we are able to collect exceptionally rich data on the HBS student characteristics. Moreover, the course setting allows us to precisely define the choice set that each student has when it comes to team formation, a relevant feature that differentiates us from the typical social or professional network setting.

Second, our paper contributes to the understanding of the causal relationship between horizontal team diversity and performance. Theoretical work on diversity focuses on the trade-off between information gains and communication costs. Heterogeneous teams benefit from more diverse pools of skill and knowledge. Still, at the same time, differences in race/ethnicity, culture, and mother language hinder efficient communication among team members, thus potentially lowering productivity (Alesina and La Ferrara 2003, Lazear 1999). With field experiments, Hoogendoorn and Praag (2012) and Hoogendoorn, Oosterbeek, and Van Praag (2013) find the benefit of information sharing is greater than communication cost in more diverse teams. Marx, Pons, and Suri (2021) find horizontal diversity in race/ethnicity decreases team performance because people in heterogeneous teams are more likely to complain about their teammates. Knippenberg and Schipper (2007) review the empirical literature on team diversity and performance from 1997 to 2005. They conclude that the empirical results on diversity are "highly inconsistent" because of the endogenous process of group formation in the majority of the existing research.

Our study provides an empirical setting to test the causal relationship between horizontal diversity and team performance more cleanly. By exploiting the quasi-random variation in team assignments for the 2013 cohort, where a computer algorithm was used to assign students to teams, we find that race/ethnicity diversity hurts team performance. In other words, exogenously homogenous teams performed better than exogenously formed diverse teams. When we look at the intersection of race/ethnicity and gender, we find that the effect of homogeneity on performance is driven by the joint homogeneity of both gender and race/ethnicity. Student teams who match on both dimensions at the same time performed the best. This result is consistent with randomly assigned diversity reducing communication efficiency and increasing the probability of conflict within the team.

Although HBS utilized a computer algorithm to assign teams randomly in 2013, they changed the assignment to be voluntary formation from 2014 to 2016. Interestingly, we find the negative performance gradient of racial and ethnic diversity in the 2013 cohort is largely alleviated in the 2014-2016 cohorts when teams are formed voluntarily by approximately 60%. Importantly, the presence of two different team assignment mechanisms, one being completely "hand-on" and one being completely "hands-off" from the administrator's perspective, is unique to our setting and rare in the literature. Concretely, the presence of both treated cohort and control cohorts allows us to identify the causal impact of the diversity intervention on the performance gradient, as potential confounders in the single cohort regression will be differenced out. The differential performance effects that arise from forced diversity created by random assignment versus endogenous diversity created by voluntary formation has important policy implications: It suggests that how a diverse team is formed has a material impact on its ability to function well together.

Finally, our paper looks at the performance effects of diversity in the vertical relationship between supervisors and team members. Athey, Avery, and Zemsky (2000) model the role that shared characteristics in the mentor-subordinate relationship play in performance and promotion. Findings from the empirical literature show a mix of negative, null, and positive effects of vertical diversity. Carrell, Page, and West (2010) find that professor gender has a large effect on female students' performance in math and science classes using the U.S. Air Force Academy data. However, Bednar and Gicheva (2014) look at NCAA Division I athletic departments and find no effect of the gender of the athletic director on female representation within coaching and staff positions. Bagues, Sylos-Labini, and Zinovyeva (2017) find that women in evaluation committees for professorships are not more favorable toward female candidates. Matsa and Miller (2011) find that boards with more female directors are associated with an increased number of women in senior management positions. On the other hand, Bertrand, Black, and Jensen (2019) do not find that mandates on gender quotas in Norwegian corporate boards had much effect on women in business. Marx et al. (2021) find that vertical diversity increases team performance, as workers tend to exert more effort when the manager is from a different ethnic background.

Exploiting the random assignment of faculty members to students, we estimate the causal impact of vertical diversity, especially gender ties, on the performance of the entrepreneurial teams. These gender ties are exogenous to the gender make-up of the team because students had no control over who their faculty members were. When the faculty section leader is female, team performance increases monotonically as the fraction of female students in a team increases. We do not find significant results among external female judges, who serve as evaluators for these teams but do not interact with the students as the faculty section leader does. Compared to the literature, the presence of both faculty section leaders and external judges in our setting allows us to identify the impact of vertical diversity separately for mentors and evaluators. The result is consistent with the importance of mentorship and the positive performance impact women experience when mentored by women.

Our results on the diversity and its performance implications are important beyond the context of our research setting. First, the main criteria for evaluating the micro-businesses were related to the actual business viability and the ability to attract real customers, judged by experienced venture capital investors and entrepreneurs. Second, a significant minority of these micro-businesses continued to operate after the semester, and many raised external funding, including venture capital. Finally, among the 3,864 MBA students in our sample, over 20% of them work in venture capital or technology-related areas after graduation, representing a sizable labor inflow to the entrepreneurial eco-system. Thus, understanding the strength of homophily in team formation sheds light on the lack of diversity in entrepreneurship and venture capital (Calder-Wang and Gompers 2017).

Further, our findings have important implications on broader policy considerations for workplace diversity initiatives. Workplace diversity, or more broadly inclusivity, has both short-term and long-term performance implications for firms and workers. Such policies generate long-term effects because education or occupation choices tend to be forward-looking and entail a significant time lag. Some policymakers have promoted diversity in the workplace by implementing gender (or race/ethnicity) quotas in recent years. For instance, the Norwegian government enforced a gender quota on corporate boards (Ahern and Dittmar 2012). More recently, California has become the first state in the US to mandate board diversity. Our results suggest that for such mandated diversity policy to bring about performance benefits, it is important to ensure match-specific qualities are considered adequately in the team formation process and to ensure existing biases against certain groups could be reduced.

2. Empirical Setting

First-year MBA students at Harvard Business School were required to take FIELD (Field Immersion Experience for Leadership Development). In the third model of this course (i.e., FIELD 3), students work in small teams to develop and launch a "microbusiness." FIELD 3 is designed to allow MBA students to "hone their collaborative skills while experiencing the challenge and excitement of being an entrepreneur." Here is a high-level overview of the course structure:

i. At the beginning of the spring semester, students form a team of 5-7 with members all from the same section.[1]

ii. Throughout the semester, these student teams work together to develop and build and their micro-businesses. They are required to develop a business idea, gather market feedback, create the product/service, manage external resources and vendors, and market and sell the product/service.

iii. After three months, on the "Launch Day," each team makes a presentation, and those that do not have a product ready to sell are moved to the "Failed Business Track" at the discretion of the faculty member leading the section.

iv. Lastly, at the end of the semester on the "IPO Day," the surviving teams proceed to present their projects to a panel of judges from the academic, corporate, entrepreneurship, and venture capital world. The judges rank the teams by determining whether they have demonstrated product demand, whether the business is viable (e.g., positive cash flow in a five-year period), and their ability to create the most value.

Importantly, FIELD 3 was taught to 4 cohorts of students using the same course structure but one critical difference in terms of the team formation mechanism: Namely, for the Class of 2013, team memberships were randomly assigned by a computer algorithm developed by the HBS administration. In contrast, for the Class of 2014, 2015, and 2016, the teams were formed voluntarily by students. For the Class of 2013, the computer algorithm randomly assigns students to teams conditional on their characteristics observed by the course administrator. Concretely, the algorithm is developed to ensure that the composition of each team created by the computer approximately reflects the overall composition of the entire section in terms of gender and whether a student is

[1] Each year, approximately 900 students matriculate at the Harvard Business School. They are divided into 10 sections of approximately 90 students each. All students take the same first-year mandatory courses with the same section-mates throughout the year.

domestic or international. In other words, the probability that student *i* matches with student *j* is the same as the probability that student *i* matches with student *j'*, conditional on *j* and *j'* having the same gender and international status.

The structure of the Harvard Business School curriculum and the set-up of the FIELD 3 course offer us a unique and powerful setting, allowing for credible identification of the impact of diversity on performance. First, for the Class of 2013, because the assignment of the team members is performed by the computer algorithm, conditional on gender and international status, the identity of one's teammates and their characteristics are independent of student characteristics or unobservable student preferences. From the student's perspective, they cannot choose the identity of their teammates.

Second, the choice of the team formation mechanism (i.e., computer assignment vs. voluntary formation) is also orthogonal to student characteristics, allowing for credible inference on the impact of such diversity intervention. We believe it is rather implausible that a sizable portion of the students defers their matriculation decisions based on the likely team assignment process for one of the three modules of a specific course out of over ten mandatory first-year courses. In other words, the random assignment of the team-formation mechanism across different cohorts allows us to identify the performance impact of the diversity intervention created by the computer algorithm relative to the control group with no intervention. Compared to typical field experiments or natural experiments that study the impact of a particular diversity intervention, such as Hoogendoorn, Oosterbeek, and Van Praag (2013) and Marx, Pons, and Suri (2021), the presence of both the treated and control cohorts in our setting is rare in the literature. Importantly, it allows for a credible inference because we can difference out the baseline impact using the control cohorts.

Third, because HBS utilizes a sectioning algorithm that divides the entire student body into ten equal-sized sections and teaching are performed in parallel across all sections, the assignment of the faculty leader to each student is also exogenous to student characteristics, allowing us to obtain credible inference on the performance implications of vertical diversity. These students have no ability whatsoever to choose which section they join. The balanced team assignment algorithm in FIELD 3 is adapted from the balanced sectioning assignment used by HBS to divide the entire incoming class into sections. Each section reflects the student composition of the entire class. Shue (2013) utilizes the randomization in the HBS sectioning algorithm to derive the causal impact of peer effects. It is important to note that neither the balanced section assignment in Shue (2013) nor the balanced team assignment in our setting create a problem for the validity of the empirical methodology: From an individual student's perspective, who their faculty section leader will be and who their team members will be for the course is exogeneous to their own preferences.

3. Data

We collect an extensive set of information from the Office of MBA Student and Academic Services at Harvard Business School. It includes anonymized student characteristics and background information, the team membership of each student, the team performance, and the characteristics of the panel of judges at the IPO Day.

Student Characteristics and Background:

We obtain anonymous data on student characteristics from the Office of MBA Student and Academic Services. We observe the gender, race/ethnicity, home country, undergraduate institution, past employers, and the industry of each MBA student from the class years 2013 to 2016. We were not provided with students' actual names. Table I reports the summary statistics for the 3,684 MBA students in our sample. Women make up 41% of the total student population. Approximately 40% of the students are white Americans, 12% are Asian Americans, 5% are Black, 4% are Latinx Americans, and 35% are international students. India, Canada, and China represent the top three origin countries for international students (Appendix Table I). In terms of past work experience, roughly half of the students worked in finance or consulting before business school. Not surprisingly, the big three consulting firms (McKinsey, Bain, and BCG) and bulge bracket investment banks (Goldman and Morgan Stanley) are the top five past employers for Harvard MBA students, shown in Table II. On average, 11% of students had experience in the technology industry. Notably, this number increased substantially by more than 50%, from 9% for the Class of 2013 to 14% for 2016. 25% of the MBA students graduated from Ivy League schools; Harvard, Stanford, and the University of Pennsylvania are the top 3 undergraduate institutions (Table II).

Team Membership and Validation of Assignment Mechanism:

We also obtain the team selection of each student. From 2013 to 2015, there were 150 teams in each class year, and the average team size was 6. In 2016, the average team size was changed to 5, and there were 180 teams.

We provide some initial graphical evidence showing that the computer assignment of teams for the Class of 2013 sought to create balanced teams in terms of gender and international status. Figure I Panel A shows the distribution of female students in each team for both team assignment mechanisms. For the Class of 2013, the distribution of the gender composition indicates that the computer algorithm created gender-balanced teams: Out of 150 six-person teams, 62% have two women, and 38% have three women. 1 team has two women. There are zero teams with more than three women. There are also zero teams with no women. In sharp contrast, when students form teams voluntarily in 2014 and 2015, out of 300 six-person teams, 12% of the teams have no women, 12 % of the teams have one woman, 53% of the teams have 2 or 3 women, and another 19% of the teams have three or more women. Figure I Panel B shows that the distribution of international students across the team assignment mechanism. While there were only two teams with no international students in 2013 when teams were randomly assigned by the algorithm, 16% had no international students in 2014 and 2015 with voluntary team formation. Similarly, only 4% of the teams had more than four international students in 2013 with computer assignment, but 20% had over four international students with voluntary formation.

Next, we provide a framework to validate the conditional random assignment used in 2013 more formally. The framework also forms the basis for identifying and quantifying the strength of homophily in team formation when students form teams voluntarily as well. Specifically, we construct student-student pairs by matching each student to every other student within the same section and year. This process creates over 335,686 potential pairs, with 81,368 potential pairs for the Class of 2013 and 254,318 for the Class of 2014-2016. We then create a dependent variable *real_match*, which equals one if the two students are members of the same team and 0 otherwise. The independent variable gender (race/ethnic, school, industry) tie equals one if two students belong to the same gender (race/ethnic, school, industry) group.[2] Our data construction method is similar to Louch (2000).

To illustrate, consider the following example: James Brown is a Section A student in 2013, and he has five teammates. We match Mr. Brown to all his section mates (89 of them) by creating 89 student-student pairs. Intuitively, each pair is a potential teammate with whom Mr. Brown could be

[2] Race/ethnicity ties are defined as follows: For domestic students, the race and ethnicity groups are Whites, Asian Americans, Latinx, and African Americans. For international students, the race and ethnicity groups instead measure the broad regions that the students come from: Europe, Latin America, South Asia, East Asia, the Middle East, and Africa. Two students share the race/ethnicity tie when they belong to the same racial or ethnic group if they are domestic, or if they hail from the same region if they are international.

paired. If the match happened randomly, Mr. Brown would pair with an arbitrary teammate with a probability of 5.6% (1 out of 89). Variable *real_match* equals 1 for the five pairs for which Mr. Brown is matched to his real teammates. To measure the effect of the computer assignment or homophily on matching, we compare the probability of matching conditional for a pair having the same gender (race/ethnicity, school, and industry) to the probability of matching for a pair with different gender (race/ethnicity, school, and industry).

Table V validates that, once conditional on gender and international status, the matching probabilities generated by the computer algorithm in 2013 are independent of shared demographic characteristics or personal backgrounds. In column (1), shared gender or shared international status negatively predicts matching probabilities, indicating that the computer algorithm sought to create balanced teams on these two characteristics. However, other coefficients such as race/ethnicity ties, school ties, and industry ties have no predictive power on whether two students will match in a team. Columns (2) to (3) show that the same patterns hold true with subsamples of student pairs that either share the same gender or belong to different gender groups. Consistently, columns (3) to (6) show that other personal backgrounds do not matter to team assignment in the subsamples of student pairs that share/do not share their international status. As described in the next section, these coefficients generated by the computer assignment stand in sharp contrast to the matching regressions for the 2014 to 2016 cohorts with voluntary formation.

Team Performance and Judge Characteristics:

Beyond student characteristics and team membership, we also collect information about the team performance from the MBA program office, summarized in Table III. Specifically, we code team outcomes into four binary indicators:

1) IPO Day: We observe whether a team progresses to the IPO Day or whether it was placed in the "Failed Business Track." Approximately 75% of the teams were determined to be sufficiently developed to present on IPO Day.
2) Viable: A team that presented on IPO Day was deemed by judges to be a viable business if they believe the business could be cash flow positive over a five-year period. Roughly 50% of all projects were deemed "Viable".
3) Section Top 3: A project was ranked in the top 3 of their section by a panel of judges. Since a typical section has approximately 15 teams, about 20% of all projects were ranked

as Section Top 3.

4) Class Top 3: A project was ranked as top 3 in the entire class (2%). Since there were 150 teams in each cohort in 2013-2015 and 180 teams in 2016, approximately 2% of all projects were ranked as Class Top 3.

Correspondingly, we construct a composite performance score based on the median of the quantile of the team's outcome. For those who did not progress to IPO Day, their performance score is set to 0.125 because 25% of the teams do not progress and the median quantile of this group is 0.125. Analogously, among those who advanced to the IPO Day, but the project was deemed not viable, their performance score is set to 0.375 because they fall between the 25th and 50th percentile of the class. Among those who had a viable project but not Section Top 3, their score is 0.65, as they fall between the 50th and 80th percentile of the class. Section Top 3 teams are assigned a score of 0.89, as they fall between the 80th and 98th percentile. Finally, Class Top 3 teams are assigned a score of 0.99, as they fall in the top 2 percentile. Our performance measure is increasing in the project outcome.

In addition, for every section in 2014-2016, we obtained the demographic characteristics (e.g., gender, race, and ethnicity) of the panel of judges at the IPO Day. Each panel comprises a faculty section leader who led instructional sessions throughout the semester and external judges from the industry, shown in Table IV. There are 21 unique faculty section leaders, out of which 5 of them are women (24%). There are 76 unique external judges who participated in the IPO Day, 26% of whom are women.

4. Homophily in Team Formation

In this section, we show that homophily along the dimension of gender, race/ethnicity, educational background, and past work experience all play a significant role in team formation when teams are formed voluntarily. We also quantify the relative strengths of different dimensions, showing that demographic ties play a larger role than ties on acquired characteristics.

Although the homophily phenomenon of "birds a feather flock together" is well-documented in both sociology (McPherson, Smith-Lovin, and Cook 2001) and economics (Jackson 2014, Bertrand and Duflo 2017), one advantage of our setting is that we can precisely quantify the relative strengths of different components of personal, because we have the ability to precisely measure the characteristics of all those in one's choice set, namely everyone in the same section. It is much better measured than typical social settings such as friendship or professional networks, where the potential choice set is challenging to define.

Using the pseudo pair specification described in the previous section, Table VI column (1) presents the regression results for matching from 2014 to 2016 when students were allowed to choose their own teams. Race/ethnicity ties increase the probability of matching by 1.4%. Given the base rate of matching is 5.6%, this represents a 25% increase from the baseline probability of randomly matching with a student from the same race/ethnicity. Similarly, we find shared gender increases the probability of matching by 1.3%, corresponding to a 23% increase relative to the baseline. Attending the same undergraduate institution increases the probability of matching by 0.85%, a 15% increase from the baseline. Having industry experience in the same sector increases the matching rate by 0.62%, an 11% increase from the baseline. All results are significant and economically meaningful. Table VI column (2) reports the regression result using the 2013 subsample. Given that teams were randomly assigned, the coefficients on race/ethnicity tie, school tie, and industry tie are statistically not different from zero. The negative matching coefficient in front of gender refects HBS's gender-balanced assignment mechanism.

Table VI column (3) tests whether homophily based on endowed demographic characteristics is stronger than homophily based on acquired characteristics. The variable *Endowed Demographic Match* is an indicator variable equaling one if race/ethnicity tie or gender tie equals one. *Acquired Characteristics Match* equals one if school tie or industry tie equals one. In the 2014-2016 subsample, the coefficient on Endowed Demographic Match is more than twice as large as the coefficient on Acquired Characteristics Match, and the difference is statistically significant at the 1% level.

The main homophily results found here are robust to alternative statistical inference procedures. In the main tables, the standard errors are clustered at the student level. Namely, every 89 pseudo pairs for every given student is considered a cluster. Such clustering procedure is commonly used in the literature (Sufi 2007, Corwin and Schults 2005). Alternatively, we still find all the results remain statistically significant using a randomization inference procedure. Specifically, we create 1,000 random permutations of alternative team formations and re-run the matching regression on the generated data. In Appendix Table III, we find that all of the actual matching coefficients belong to the extreme right tail of the simulated coefficients, admitting a p-value smaller than 5%. Moreover, we find that the standard errors of the simulated coefficients fall closely in line with the analytical standard errors (e.g., 0.0012 vs. 0.0011 for the race/ethnicity tie standard errors), suggesting that asymptotic normality is a reasonable assumption for our setting.

4.1 Breakdown of Homophily by Subgroups

Because we have detailed demographic and personal backgrounds of all students, we could further investigate the relative strengths of homophily by subgroups.

By Race / Ethnicity / Nationality:

Along racial and ethnic lines, the propensity to match is strongest among international students hailing from the same region. Among domestic students, we find that both Asian Americans and White Americans have strong tendencies to form teams within their group. Along gender lines, we find that women are more likely to match within the same gender than their male counterparts.

Table VII, column (1) shows that the probability of matching based upon shared race/ethnicity increases by 1.2% and 1.4% among White American and Asian American MBA students, respectively, translating to over 20% increase relative to the baseline of random matching. The coefficient for African American students is 1.3 %, but we lack statistical power likely because African American students only make up 5% of the student body. Latinx Americans seem no more likely to match with other Latinx Americans.

The propensity to match is highest among international students from the same region. An international MBA student is 4.0% more likely to find a teammate from the same region, three times greater than the effect among White and Asian Americans. A detailed breakdown of international students by region in Appendix Table IV shows that the increase is highest among students from East Asia, the Middle East, and Latin America. The coefficients for these groups are around 6%, over 100% larger relative to the baseline and twice as large as the coefficients for Europeans and South Asian students.

By Gender:

Table VII, column (2) shows that shared gender increases the probability of matching by 1.2% and 1.7% for men and women, respectively. The difference between these two coefficients is statistically significant at 1% level, suggesting women are more likely to choose a teammate with the gender than men. The negative gender coefficients for 2013 reflect that the computer algorithms used sought to create gender-balanced teams, compared to the truly random baseline.

By Education and Past Experience:

In Table VIII, we examine the effect of education ties and industry ties on matching in the student teams. In column (1), education ties are much stronger among students from non-Ivy

League schools. Specifically, attending the same non-Ivy college increases the matching probability by 1.9%. In column (2), we break down the industry ties by industry sectors. We find the effect strongest among students who worked in non-finance, non-consulting, and non-technology industries, increasing the matching rate by 2.2%. The magnitude of the effects is similar among finance, technology, and consulting industries, which is around 0.45%. Overall, we do not attempt to provide a sociological theory as to why certain subgroups exhibit stronger homophilic tendencies than other subgroups. Still, we quantify the prominent patterns that are present in the data.

4.2 Gender Differences in Homophily

While the subgroup analysis revealed that women exhibit stronger homophily towards other women, we provide some evidence that men exhibit stronger homophilic tendencies than women for all other non-gender-related demographic and personal characteristics, including race/ethnicity, education, and past experience. In Table IX columns (1) and (2), we find that the coefficients on race/ethnicity, school, and industry tie are statistically larger for men than women. We do not observe women match with section-mates based on education or industry background at all.

These results contrast with Brashears (2008), where homophily in education level is uniform among males and females using 1985 general social survey data. On the other hand, interestingly, the results here are consistent with the patterns found in real-world venture capital settings. Using the deal-level data from Calder-Wang and Gompers (2021), in Appendix Table VI Panel A, we found that women venture capital partners are much more likely to invest in women entrepreneurs. However, Panel B shows that women VCs are *less* likely to match on the race and ethnicity of the entrepreneurs compared to their male venture capital partner colleagues. Men in venture capital are more likely to invest in entrepreneurs with shared ethnic ties across all ethnicities. Relatedly, similar patterns are also documented among Wall Street analysts in Fang and Huang (2017).

Overall, this section establishes two main results that enrich the existing literature on homophily. First, we estimate the strength of homophily in gender, ethnicity, education, and industry background. We find homophily exists among both endowed demographic characteristics and acquired characteristics, but it remains substantially stronger among endowed demographic characteristics. To the best of our knowledge, our study is the first attempt to estimate and compare the relative strength of various characteristics ties in an entrepreneurial setting. Second, we examine the gender difference in homophily, an often-overlooked aspect in prior studies. We document that men exhibit a stronger tendency to match with peers with the same ethnicity/race, education, and industry backgrounds than women in our entrepreneurial class setting. We also find similar results in real-world venture capital deals where men are more likely to invest in entrepreneurs with shared racial/ethnic backgrounds than women VCs.

5. Performance Implications of Horizontal Diversity

In this section, we analyze the impact of team diversity on performance. We provide a causal estimate of the impact of diversity intervention on performance. We find that increased team diversity created by random assignment is harmful to performance. However, we find that diverse teams formed voluntarily exhibit a much weaker penalty on performance. Overall, the finding highlights how diversity is created in an organization is instrumental in whether it will generate better outcomes.

5.1 Definition of Diversity Scores

Our unit of performance analysis is at the team level. There are 150-180 teams in each class year, and each team has 5-7 students. We measure team diversity across four different dimensions: race/ethnicity, gender, school, and industry, and construct the diversity measure for each dimension as the following:

$$Diversity\ Score\ _i = 1 - \frac{\#\ of\ ties\ between\ team\ members\ in\ team_i\ with\ \ shared\ characterisitcs}{Total\ possible\ ties\ in\ team_i}$$

Or equivalently:

$$Diversity\ Score\ _i = \frac{\#\ of\ ties\ between\ team\ members\ in\ team_i\ with\ \ different\ characterisitcs}{Total\ possible\ ties\ in\ the\ team_i}$$

To illustrate our race/ethnicity diversity score, consider a team with six people: Three are White, two are Asian Americans, and one is an international student from Latin America. Race/Ethnicity Score in this team will be 1-(3+1)/(5+4+3+2+1)=11/15, as there are three ties among three White team members, one tie between two Asian American students, and fifteen possible ties among all six team members. Gender Score, Education Score, and Industry Score are constructed analogously. Diversity is monotonically increasing in the score. It equals zero if everyone in the team is the same type and equals one if everyone has different characteristics.

Figure II plots the distribution of the race/ethnic diversity scores across different team assignment mechanisms. Panel A plots the probability distribution under the 2013 conditional random assignment compared with voluntary team formation. Notice that the score distribution under voluntary formation has a greater mass among lower-diversity score teams; namely, there are more homogenous teams. When plotted as a cumulative distribution function, Figure II Panel C implies a larger area under the curve for voluntarily formed teams than randomly assigned teams.

The average diversity score on race/ethnicity decreased from 0.76 under random assignment in 2013 to 0.72 under voluntary formation in 2014-2016. The average diversity score on gender also decreased from 0.56 for teams created under random assignment to 0.43 for teams formed voluntarily. The results above are consistent with stronger homophily with voluntary team formation documented in the previous section.

5.2 The Impact of Diversity on Performance

As described in the data section, we use the median quantile of the team's project ranking as our performance measure. In this section, we examine the impact of diversity on team performance.

Graphically, Figure III shows the binscatter of team performance on diversity scores. Panel A on the left indicates that among randomly assigned teams (2013), higher diversity scores correspond to poorer team performance. Panel B on the right shows that among voluntarily formed teams (2014-2016), the correlation between performance and diversity is much smaller, with the slope of the binscatter much flatter than the left panel.

Table X column (1) shows the coefficient on the race/ethnicity score is -0.49 for the 2013 cohort with randomly assigned teams. Since the standard deviation of the race/ethnicity score is 0.18[3], it suggests that a one-standard-deviation increase in racial/ethnic diversity leads to approximately 9 percentiles decline in the performance rankings (e.g., a decline from being ranked at 80th percentile to 71st percentile). In other words, in 2013, where teams were exogenously assigned, relatively more homogenous teams performed better than more diverse teams in terms of race/ethnicity.

Because the assignment of teams is random in 2013, these students have no ability whatsoever to select teams based on unobservable student quality or preference. Thus, the negative relationship between team diversity and team performance in 2013 admits a causal interpretation: Higher racial and ethnic diversity levels lead to worse team performance in our entrepreneurial team setting.

[3] Because the distribution of the diversity scores will vary by team assignment mechanism, we calculate the "true" standard deviation using 1,000 simulations of team formations under pure, unconditional random assignment.

The usual challenge associated with interpreting the correlation between diversity and performance causally concerns the presence of confounders. In a typical organizational environment, one may be concerned that more diverse firms are better managed and thus could attract unobservably higher-quality candidates. More diverse firms may also attract candidates who are unobservably better at collaborating with colleagues from different backgrounds. Therefore, such selection on unobservables likely leads to better firm performances, but such performance lift cannot be directly attributed to firm diversity. To address such concerns, researchers have used either field experiments (e.g., Hoogendoorn et al. 2013, Marx et al. 2021) or natural experiments (e.g., Calder-Wang and Gompers 2021) to create exogenous shocks to diversity. One key advantage of our current setting is that the conditional randomized team assignment used in 2013 is akin to a field experiment run by the HBS administration: Whether a given student will land in a diverse team or homogenous team is completely exogenous to their own unobservable quality or preference.

Although the random assignment successfully removes the scope under which the team diversity may be subject to unobserved student selection, since we do not observe the same student across different assignment mechanisms, another potential threat to our regression specification is that the diversity score may itself be correlated with other student characteristics that are directly predictive of performance. For instance, more diverse teams may have a higher fraction of Asian Americans because the diversity score is calculated using student race and ethnicity. Meanwhile, we know that in the data, Asian Americans are more likely to have experience in the technology section (Appendix Table VIII), which may be predictive of their performance in our entrepreneurship class. In other words, the presence of performance-relevant student characteristics that are also correlated with our diversity score can potentially create bias to our estimate.

To address this concern, in Table X column (2), we directly control for characteristics such as the percentage of students with startup experience and the percentage of top college graduates, two proxies for students' abilities. We observe results remain significant, and the magnitude of the coefficient stays similar (0.49 vs. 0.45). Moreover, Appendix Table VII shows that the main results also remain mostly unchanged after adding a variety of other plausible controls, including the fraction of White Americans in the team, the fraction of native English speakers, the fraction of students with work experience in consulting, finance, or technology. We recognize that there will always be a limit as to what we can control for explicitly. Nonetheless, the fact that none of the plausible performance-relevant controls we considered had a substantial impact on the results gave us more confidence in the main specification.

Lastly, Table X column (1) shows no statistically significant coefficients in front of the gender score. It is expected because the 2013 assignment algorithm creates gender-balanced teams, leaving little residual variation in the gender diversity of the team for us to exploit. Additionally, we find some evidence of the negative impact of school diversity on performance, albeit with somewhat weaker statistical significance.

5.3 The Impact of "Forced" Diversity on the Performance Gradient

We cannot interpret the correlation between diversity and performance causally for the cohorts formed voluntarily (2014-2016) because of the likely selection on unobservables. However, the presence of these control cohorts provides us the ability to identify the causal impact of the intervention itself (namely, the conditional random assignment utilized by the HBS administration relative to the voluntary team formation) on the performance gradient (e.g., the slope of performance on diversity). Quantitatively, when HBS changes the way it creates team diversity from random assignment to voluntary formation, the negative performance implication of diversity on performance becomes substantially reduced by about 60%.

In Table X, comparing the results from the voluntary formation years in column (3) with the random assignment results in column (1), the coefficient in front of the race/ethnicity score is much smaller (-0.18 vs. -0.49). Although the -0.18 does not admit a causal interpretation, we could run a full-sample regression where the difference in these two coefficients gives us the causal impact of the diversity intervention on the performance gradient. In column (5), the coefficient on the interaction term Voluntarily Formed × Race/Ethnicity Score is 0.29, suggesting that changing the team assignment from random assignment to voluntary formation reduces the negative impact of diversity on performance by approximately 60% (0.29/0.49). However, it is also at the cost of creating fewer diverse teams overall.

Our unique empirical setting allows us to interpret the impact of diversity intervention (in the form of random assignment) on the difference in the performance gradient causally. The relevant exogeneity condition is that the treatment assignment (i.e., random team assignment) is orthogonal to any unobservables that would cause the performance gradient to be different. We believe that it is rather implausible that the performance gradient on diversity may fluctuate from cohort to cohort from 2013-2016 in a way that systematically favors the 2013 cohort for some reason. Because the admission process and the school curriculum from 2013 to 2016 have remained stable, the use of random assignment in 2013 is independent of any unobservables that may affect the counterfactual performance gradient for the 2013 cohort.

In addition, the causal interpretation of the change in performance gradient is not subject to the confounders mentioned above that could bias the performance gradient estimate for the 2013 cohort. To the extent that there may be omitted performance-relevant student characteristics (e.g., the fraction with startup experience, which is correlated with both performance and the diversity score), such bias is likely stable across these cohorts. Thus, it is differenced out in the full sample specification.

After all, the presence of two different types of team assignment mechanisms is rare in the literature of field experiments. The ability to compare different assignment mechanisms suggests that not all types of diversity are created equal. How team diversity comes about and the type of intervention used to achieve such diversity significantly impact the performance gradient. In some sense, the random assignment utilized by the HBS program may be the most extreme version of diversity intervention where no information about match-specific qualities is considered at all. Consequently, the outcome of such diversity intervention is that the performance gradient on diversity becomes much more negative than a voluntary team formation baseline. To the extent we consider the diversity created by random assignment is "forced," these results suggest that voluntary team formation alleviates as much as 60% of the underperformance of forced diversity.

5.4 Additional Properties of the Performance Gradient

In this section, we explore a few additional properties of the negative performance gradient on diversity. Although we currently cannot isolate the mechanism of the level or the change precisely, some of these findings could still shed light on the potential sources of the negative impact. We also provide several alternative specifications to ensure that the main results are robust.

First, we examine the intersection of gender and ethnicity. Given the negative coefficients on the race/ethnicity score, we find that the performance lift from team homogeneity is driven by the quadrant where students match on *both* gender *and* race/ethnicities. Concretely, we now define a composite race/ethnicity-gender diversity score, where a student pair is considered to share the same characteristics when they share the same gender and the same race/ethnicity. In other words, the score is the lowest when the student pair matches on both dimensions. Table XI column (1) shows that once we include the race/ethnicity-gender diversity score, the magnitude on the singledimensioned race/ethnicity score becomes statistically not different from zero. The coefficient in front of the composite race/ethnicity-gender score is large and significant at -0.82. With a standard deviation of 0.12, a one-standard-deviation change in this composite score translates into a change of 10 percentile in the project performance ranking. In column (2), we find that the intersection effect is similar for both men and women, although we have better statistical power for men. Columns (5) and (6) replicate our previous findings on the impact of forced diversity, where voluntary team formation largely alleviates the performance penalty of forced diversity. Overall, this analysis suggests that looking at multiple aspects of diversity at the same time may be important for understanding performance implications.

Second, we also examine the contribution of different racial/ethnic subgroups. In particular, given that there is a significant portion of international students, one may be concerned that the negative performance implication is entirely driven by the friction caused by working with students from outside of the United States. Table XII shows that the coefficient in front of the international student score, defined as the fraction of student pairs who do not come from the same region of the globe in a team, is indeed negative, significant, and large. Nonetheless, the coefficients on the diversity score on other racial and ethnic subgroups such as White American score and Asian American Score remain negative and statistically significant. We do not find a statistically significant coefficient in front of the African American score, indicating that the fraction of African American and non-African American pairs does not affect team performance. It suggests that there is limited scope for traditional taste-based discrimination to be at play in our setting. However, this may be partly because African American students only account for a small fraction (5%) of the student body.

Even though we use the median percentile as the performance measure, we could also perform our analysis using the binary outcomes that indicate whether a team's performance is above a certain threshold, namely, IPO Day, Viable, Section Top 3, Class Top 3. Appendix Table IX shows that diversity hampers performance at every stage of the project progression for the 2013 cohort. On the left tail, 75% of the teams progress to IPO Day, where more homogenous teams are more likely to make it. On the right tail, only 2% of the teams are considered Class Top 3, where the most homogenous teams in terms of alignment in both race/ethnicity and gender are most likely to be ranked at the top. Consistent with our main results, such negative performance implications are also greatly alleviated across all levels of outcomes when teams are formed voluntarily.

Given the underlying distribution of the diversity score is skewed to the right, as shown in Figure II, we also provide a robustness test where the diversity score is measured in terms of its percentile in the underlying distribution of diversity scores generated with an unconditional random assignment. In other words, if there were no homophily, the diversity score percentile would admit a perfectly uniform distribution between 0 and 1. To the extent that there may be non-linearity in the raw diversity score, the variable transformation from diversity scores to diversity percentiles addresses this concern. We show in the appendix that all main results on the negative performance impact and the improvement in the performance gradient due to voluntary formation remain robust under the alternative measure of diversity.

Overall, in this section, we provide credible inferences that diversity negatively impacts performance, leveraging the unique empirical setting with the FIELD 3 course. We also find that the performance penalty of forced diversity such as those created by random assignment becomes greatly alleviated when diversity is created with voluntary team formation. This result has important implications for policies that use gender/ethnicity quotas to promote diversity. Our results suggest that it may be essential to consider match-specific qualities beyond observable demographic characteristics in fostering a well-functioning diverse team.

6. Performance Implications of Vertical Diversity

The role of diversity has been examined in the literature at a horizontal level (i.e., among team members). It has also been examined vertically (i.e., between a supervisor and their subordinates). In our setting, we explore the relevant vertical relationships in FIELD 3, namely, the faculty section leader's role and external judges' role. Specifically, we examine whether a greater overlap between student attributes and the attributes of their faculty section leader and/or judges may influence team performance.

One crucial advantage of our empirical setting is the Harvard Business School's randomized sectioning algorithm allows us to obtain credible causal inferences on the performance impact of vertical diversity. Every year, HBS randomizes every incoming class of over 900 students into ten equal-sized sections, where students in each section sit together to take their entire first-year mandatory courses. These mandatory first-year classes, including FIELD 3, are taught in parallel with the same curriculum content but by different faculty members. Thus, from the student's perspective, the demographic characteristics of their faculty section leader and the external judges are exogenous to their own characteristics.

The outcomes for each team were determined through the development of a micro-business and evaluation of those businesses by a panel of judges on IPO Day. The faculty section leader is a member of the HBS faculty who supervised the section over the entire semester. The role of the section leader was critical to the team's performance because of their role in teaching and advising students throughout the period. Each panel then had an additional four or five external judges from the industry. Because of the different roles played by the faculty section leader, we analyzed the attributes of the section leader and the external judges separately.

Table IV reports the summary statistics on section leaders and judges' gender and race/ethnicity from 2014 to 2016. We were not able to obtain data from HBS on section leaders and judges for the class of 2013. Among the ten section leaders in each class year, there were three women in 2014 and 2015 and two women in 2016. The majority of the faculty section leaders are White, with a few Black, Latinx, and South Asians. There were no East Asian section leaders in our sample. There were more than 40 judges in each class year in our sample. The percentage of female judges increased from 14% in 2014 to 34% in 2016. The percentage of ethnic minority judges varied between 5% and 10% for each subgroup. Because there were so few minority judges, we focus on the gender ties between the section leader or judges and the students.

In Figure IV, we sort all teams into four quartiles based on the percentage of female team members and plot their performances. Conditional on having a female section leader, Panel A shows that team performance increases monotonically as the percentage of women on the team increases. In these sections, the percentages of teams progressing to the IPO day, being rated as viable, and being ranked section top 3 are 53%, 28%, and 8%, respectively, for teams with a low fraction of members (Quartile 1). These numbers increase to 90%, 76%, and 38%, respectively, for teams with a high female percentage (Quartile 4). The economic magnitude of performance increase is significant. For instance, teams with the highest number of female members were four times more likely to be in the section top 3 than teams with few or no female members when the section leader was a female. However, we do not find any relationship between male section leaders and the performance of male-dominated teams. Figure IV Panel B shows that team performance does not vary with the percentage of women (or men) in the team in sections with male section leaders. The results offer some suggestive evidence that women's performance may be improved with women mentors and supervisors.

Because the faculty section leaders serve both as a mentor and an evaluator to the teams, to separately parse out the impact of the evaluator, Table XIII presents the regression results for performance conditional on shared gender attributes between the teams and the section leader as well as external judges. The dependent variable is team performance. The key independent variable is the interaction term between being a female section leader/judge and the percentage of women on a team. Consistent with Figure IV, the coefficient on the interaction term on section leaders is positive and statistically significant at the 1% level, indicating teams with a greater number of female members perform significantly better in sections with female section leaders. The standard errors are clustered at the section level. The results remain robust if we cluster the results at the faculty level since a few faculty members taught multiple sections across different years.

Meanwhile, when it comes to external judges, we do not find any performance impact of female judges on teams with a greater number of female team members. In the second column of Table XIII, we regress performance on the interaction term between Have Female Judge and the percentage of females on the team. Have Female Judge is a dummy variable that equals one if at least one female judge is on the panel. The coefficient on the interaction term is positive but not statistically significant. The magnitude of the coefficient is also much smaller.

Overall, taken together, we believe that our result does not indicate a more favorable (or stringent) ranking of teams with female team members by female evaluators (e.g., Bagues, Sylos-Labini, Zinovyeva 2017). Instead, the likely channel is that female section leaders may have provided better mentorship throughout the year and during the Field 3 course for female students.

7. Conclusion

In this paper, we leverage various sources of randomization unique in our empirical setting to study the impact of homophily on entrepreneurial team formation and the effects of diversity on performance. We quantify the relative strengths of homophily in gender, ethnicity, education, and industry background in entrepreneurial team formation. We find that the effect of endowed demographic attributes (gender and race/ethnicity) is much stronger than team choice based upon acquired characteristics (education and industry). We also find interesting gender differences in homophily. Men exhibit a stronger tendency to match with peers with the same ethnicity/race, education, and industry backgrounds than women. Then, when we examine the effect of horizontal team diversity on performance, we find that teams in the 2013 cohort for which team membership was exogenously assigned, greater diversity across race/ethnicity led to poorer performance than more homogenous teams. When team formation was endogenous, however, such underperformance was much alleviated. Lastly, in terms of vertical diversity, we find that shared gender ties between female team members and their female faculty section leaders enhance team performance.

Our results have important real-world implications. Because the goal of the course was to form real micro-businesses, the performance was evaluated by a panel of experienced venture capital investors and entrepreneurs from the industry. In fact, a significant minority of businesses started during FIELD 3 continued to operate after the course, with some attracting significant outside funding. Moreover, many of these MBA students have chosen to work in startups and the venture capital industry after graduation. Based on the exit surveys, approximately 20% to 25% of the graduating class entered into the technology sector or venture capital during this period. Many of them later on progress to leadership positions in their field. It is reasonable to infer that such homophily found in our setting also exists in startup team formation, venture capital investing, and hiring.

Our results on the performance effects of horizontal diversity highlight the need to design and implement diversity policies thoughtfully. Although the conditional random assignment implemented for the 2013 cohort may be thought of as a draconian way to create balanced teams, it exposes the potential harm as we find a strong negative relationship between diversity and performance among these teams. The fact that much of the negative performance gradient is alleviated with voluntary team formation suggests that students could match on other characteristics that are not used by the computer algorithm, which in turn dampens the negative effect of diversity. One limitation with our data is that we are not able to extract the exact pieces of information unobservable to the computer algorithm that has led to the performance improvement under voluntary formation. One may hypothesize that it may include shared career or personal interests, working styles, risk preferences, etc. Despite the limitation, we offer a lower bound and an upper bound on the performance gradient with respect to two extremes scenarios: We allow for *no* selection on unobservables using the computer assignment in 2013, and we allow for selection on all unobservables using voluntary formation in 2014-2016. The results suggest that diversity interventions may create unintended negative outcomes unless adequate considerations about these match-specific qualities are considered.

In addition, to ensure the benefits of diversity in entrepreneurship, one needs to think carefully about how subtle treatment effects may dislodge existing biases. To harness the full benefits of diversity, policymakers need to eliminate bias against underrepresented groups. For instance, Calder-Wang and Gompers (2021) show that when venture capitalists have more daughters, they are more likely to hire a female investor, and subsequent firm performances improve after hiring.

Our results for the performance effects of vertical diversity have potentially important implications for female-led startups. The relationship between female teams and female section leaders in our setting resembles the relationship between female entrepreneurs and female VCs. Calder-Wang and Gompers (2021) and Gompers et al. (2020) document that females VC (and entrepreneurs) are underrepresented and under-supported. An effective policy to help women succeed in entrepreneurship and venture capital needs to take advantage of the superior mentorship that female venture capitalists may be able to provide to female entrepreneurs. It argues for increasing the number of women in venture capital as a prerequisite for greater representation and performance of female entrepreneurs.

Reference

Ahern, K. R., & Dittmar, A. K. (2012). The changing of the boards: The impact on firm valuation of mandated female board representation. *Quarterly Journal of Economics*, 127(1), 137-197.

Alesina, A., & Ferrara, E. L. (2005). Ethnic diversity and economic performance. *Journal of Economic LOiterature*, 43(3), 762-800.

Athey S., Avery, C., & Zemsky, P. (2000). Mentoring and diversity. *American Economic Review 90*(4), 765-786.

Bagues, M., Sylos-Labini, M., Zinovyeva, N. (2017) Does the Gender Composition of Scientific Committees Matter? *American Economic Review* 107(4): 1207-1238.

Becker, G.S. (1957). *The Economics of Discrimination."* Chicago: Chicago University Press.

Beckman, C. M., Burton, M. D., & O'Reilly, C. (2007). Early teams: The impact of team demography on VC financing and going public. *Journal of Business Venturing, 22*(2), 147-173.

Bednar, S., Gicheva, D. (2014). Are female supervisors more female-friendly. *American Economic Review, 104*(5), 370-375.

Bertrand, M, Black, S., Jensen, S. (2019). Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway. *Review of Economic Studies*, 86, pp. 191-239.

Bertrand, M. and Duflo, E. (2017). Field Experiments on Discrimination. *Handbook of Economic Field Experiments,* 1, pp.309-393.

Block, P., & Grund, T. (2014). Multidimensional homophily in friendship networks. *Network Science, 2*(02), 189-212.

Brashears, M. E. (2008). Gender and homophily: Differences in male and female association in Blau space. *Social Science Research*, *37*(2), 400-415.

Calder-Wang, S.Q. & Gompers, P. A. (2018a). Diversity in innovation. Harvard Business School Working Paper, No. 17-067, January 2018.

Calder-Wang, S.Q. & Gompers, P. A. (2021). And the children shall lead: The effect of children's gender on hiring and performance in venture capital. *Journal of Financial Economics,* Vol 142 (1), pp. 1-22.

Carrell, S., Page, M., and West, J. (2010). Sex and Science: How Professor Gender Perpetuates the Gender Gap. *The Quarterly Journal of Economics*, August, pp. 1101-1144.

Corwin, S.A. and Schultz, P. (2005). The role of IPO underwriting syndicates: Pricing, information production, and underwriter competition. *The Journal of Finance*, 60(1), pp.443-486.

Chen, R.,& Chen, Y. (2011). The potential of social identity for equilibrium selection. *American*

Economic Review, 101(6), 2562-2589.

Chen, Y., & Li, S.X. (2009). Group identity and social preferences. *American Economic Review, 99*(1) 431-457.

Clogg, C. C., Petkova, E., & Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. American Journal of Sociology, 100(5), 1261-1293.

Cohen, L., Frazzini, A., & Malloy, C. (2010). Sell-side school ties. *Journal of Finance*, 65(4), 1409- 1437.

Cornell, B. and Welch, I., 1996. Culture, information, and screening discrimination. *Journal of Political Economy*, 104(3), 542-571.

Currarini, S., Jackson, M. O., & Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4), 1003-1045.

Dobbin, F., & Kalev, A. (2016). Why diversity programs fail. *Harvard Business Review*, 94(7), 14.

Dobbin, F., Kalev, A., & Kelly, E. (2007). Diversity management in corporate America. *Contexts*, 6(4), 21-27.

Fang, L.H. and Huang, S., 2017. Gender and connections among Wall Street analysts. *The Review of Financial Studies*, 30(9), pp.3305-3335.

Fiore, A. T., & Donath, J. S. (2005, April). Homophily in online dating: when do you like someone like yourself?. *In CHI'05 Extended Abstracts on Human Factors in Computing Systems* (pp. 1371-1374). ACM.

Gompers, P. A., Mukharlyamov, V., & Xuan, Y. (2016). The cost of friendship. *Journal of Financial Economics*, 119(3), 626-644.

Gompers, P. A., Mukharlyamov, V., Weisburst, E., & Xuan, Y. (2020). Gender effects in venture capital. Forthcoming *Journal of Finance and Quantitative Analysis.*

Hampton, K. N., & Wellman, B. (2000). Examining community in the digital neighborhood: early results from Canada's wired suburb. In *Digital cities* (pp. 194-208). Springer Berlin Heidelberg.

Hoogendoorn, S., Oosterbeek, H., & Van Praag, M. (2013). The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management Science, 59*(7), 1514-1528.

Hoogendoorn, S., & Van Praag, M. (2012). Ethnic diversity and team performance: a field experiment. *Working Paper.*

Jackson, M. O. (2014). Networks in the understanding of economic behaviors. *Journal of Economic Perspectives* 28, 3–22.

Jackson, S. E., Joshi, A., & Erhardt, N. L. (2003). Recent research on team and organizational diversity: SWOT analysis and implications. *Journal of Management*, 29(6), 801-830.

Jones, S. R., & McEwen, M. K. (2000). A conceptual model of multiple dimensions of identity. *Journal of college student development*, 41(4), 405-414.

Kalmijn M. 1998. Intermarriage and homogamy: causes, patterns, and trends. *Annu. Rev. Sociol.* 24:395–421

Kets, W., & Sandroni, A. (2016). A belief-based theory of homophily. Working Paper.

Kleinbaum, A. M., Stuart, T. E., & Tushman, M. L. (2013). Discretion within constraint: Homophily and structure in a formal organization. *Organization Science, 24*(5), 1316-1336.

Lau, D. C., & Murnighan, J. K. (1998). Demographic diversity and faultlines: The compositional dynamics of organizational groups. *Academy of Management Review*, 23(2), 325-340.

Lazear, E. P. (1999). Globalisation and the market for teammates. *The Economic Journal*, 109(454), 15-40.

Lazarsfeld, P. F., & Merton, R. K. (1954). Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society*, 18(1), 18-66.

Louch, H. (2000). Personal network integration: transitivity and homophily in strong-tie relations. *Social networks*, *22*(1), 45-64.

Marsden, P. V. (1987). Core discussion networks of Americans. *American sociological review*, 122-131. Marsden, P. V. (1988). Homogeneity in confiding relations. *Social networks*, *10*(1), 57-76.

Marx, B., Pons, V., & Suri, T. (2021). Diversity and Team Performance in a Kenyan Organization. *Journal of Public Economics*, 197, 104332

Matsa, D.A., & Miller, A.R. (2011). *American Economic Review: Papers and Proceedings 101*(3), 635-639

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, *27*(1), 415-444.

Mello, A.S., & Ruckes, M.E., (2006). Team composition. *Journal of Business 79*(3), 1019-1039.

Morgan, J., and Vardy, F. (2009). Diversity in the workplace. *American Economic Review 99*(1), 472-485.

Neckerman, H. J. (1996). The stability of social groups in childhood and adolescence: the role of the classroom social environment. *Social Development*, *5*(2), 131-145.

Reagans, R. (2011). Close encounters: Analyzing how social similarity and propinquity contribute to strong network connections. *Organization Science, 22*(4), 835-849.

Reynolds, A. L., & Pope, R. L. (1991). The complexities of diversity: Exploring multiple oppressions. *Journal of Counseling & Development*, 70(1), 174-180.

Robinson, T. (1993). The intersections of gender, class, race, and culture: On seeing clients whole. *Journal of Multicultural Counseling and Development*, 21(1), 50-58.

Ruef, M., Aldrich, H. E., & Carter, N. M. (2003). The structure of founding teams: Homophily, strong ties, and isolation among US entrepreneurs. *American sociological review*, 195-222.

Sorenson, O., & Stuart, T. E. (2001). Syndication networks and the spatial distribution of venture capital investments 1. *American journal of sociology*, 106(6), 1546-1588.

Sufi, A. (2007). Information asymmetry and financing arrangements: Evidence from syndicated loans. *The Journal of Finance*, 62(2), pp.629-668.

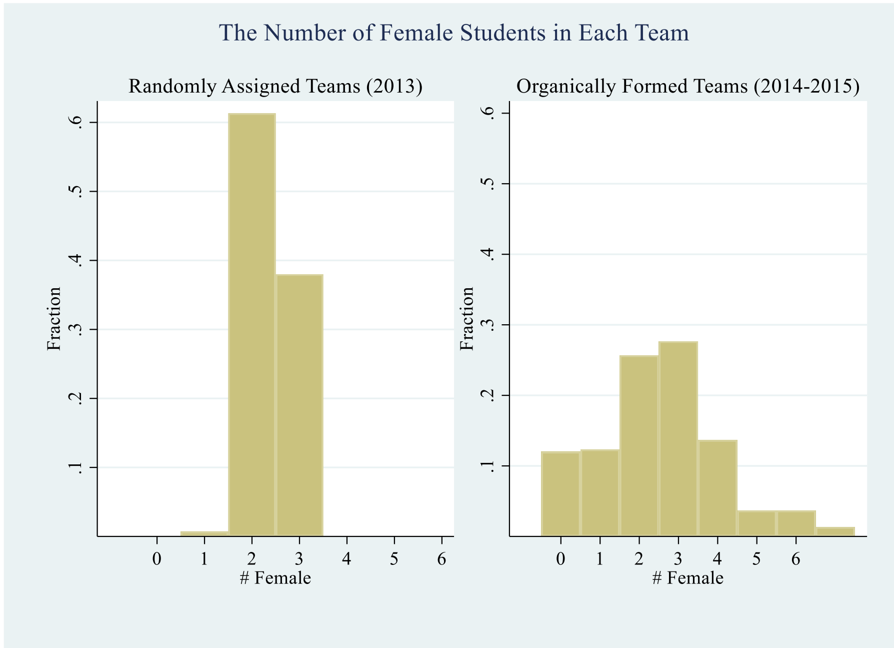
Van Knippenberg, D., & Schippers, M. C. (2007). Work group diversity. *Annu. Rev. Psychol.*, 58, 515-541.

Verbrugge, L. M. (1977). The structure of adult friendship choices. *Social forces*, 576-597.

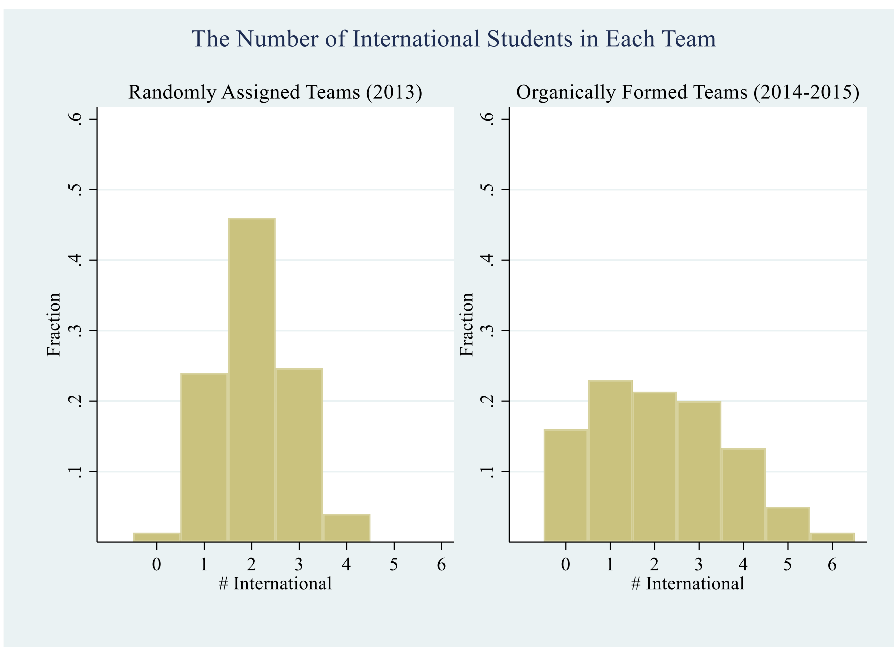
Wimmer, A., & Lewis, K. (2010). Beyond and below racial homophily: ERG models of a friendship network documented on Facebook. *American Journal of Sociology*, 116(2), 583-642.

Figures and Tables

Figure I. Distribution of Student Characteristics across Team Assignment Mechanisms

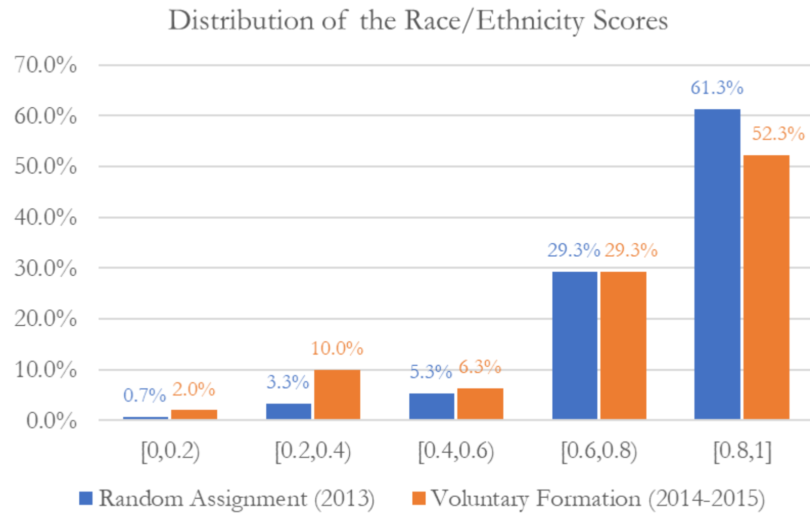


Panel A. The Distribution of Female Student Counts under Computer Assignment vs. Voluntary Formation

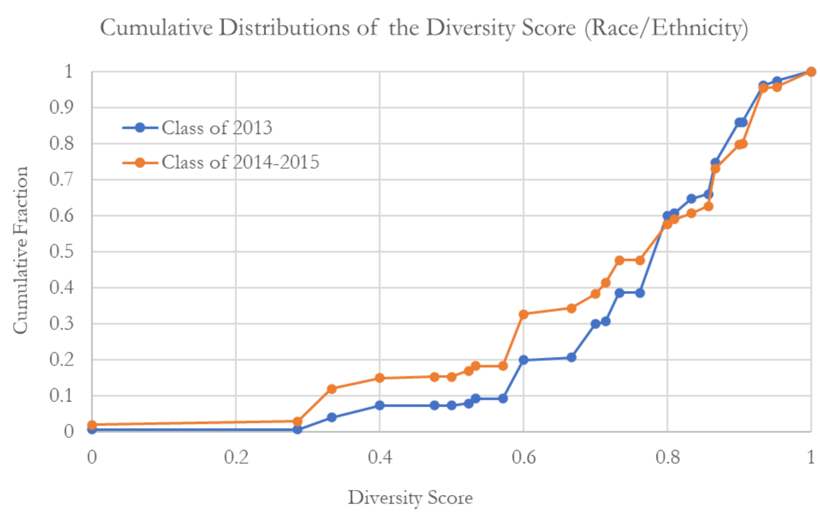


Panel B. The Distribution of International Student Counts under Computer Assignment vs. Voluntary Formation

Figure II. Distribution of Diversity Scores across Team Assignment Mechanisms



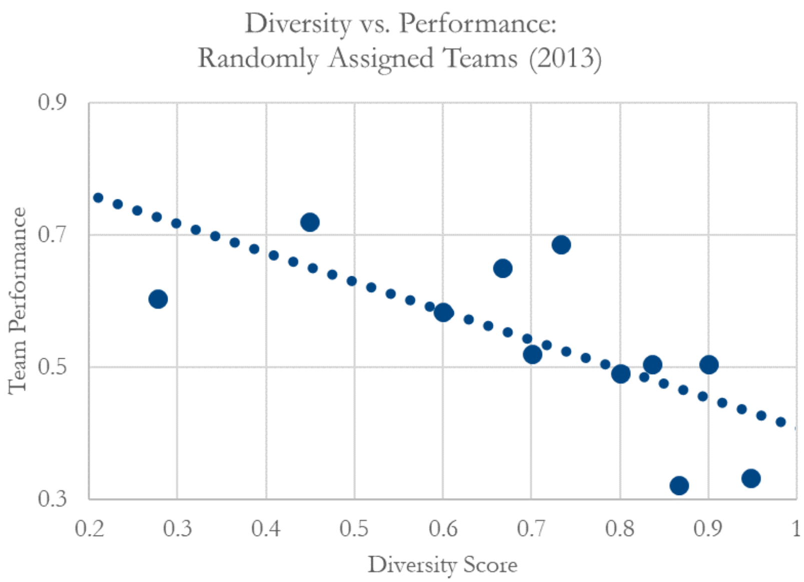
Panel A. The PDF of Race/Ethnicity Scores under Random Assignment vs. Voluntary Formation



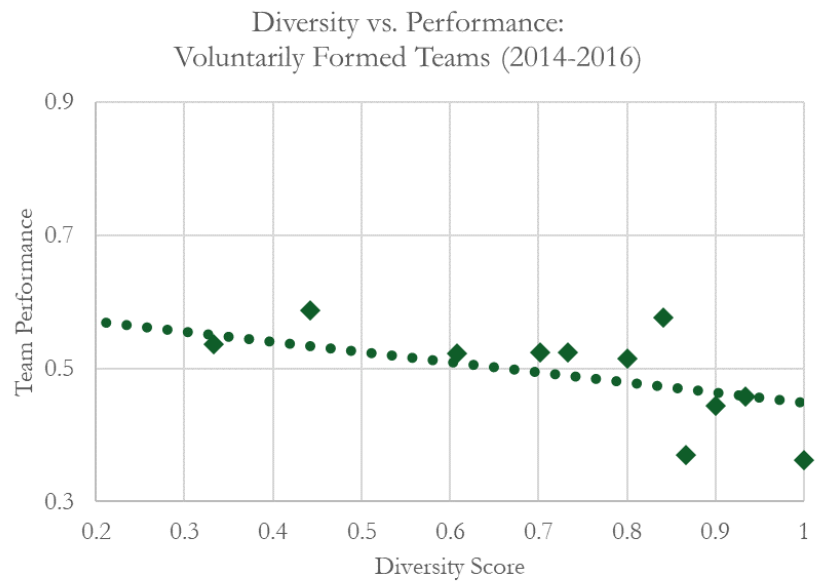
Panel B. The CDF of Race/Ethnicity Scores under Random Assignment vs. Voluntary Formation

Figure III. Horizontal Team Diversity and Team Performance

The figures below plot the binscatter of team performance by race/ethnicity diversity scores. The Y Axis is the team performance, measured as the median of the quantile of the team's outcome. The X Axis is the race/ethnicity score of the team. Larger scores imply a more diverse team. The left panel plots teams assigned randomly in 2013; the right panel plots teams formed voluntarily in 2014-2016.



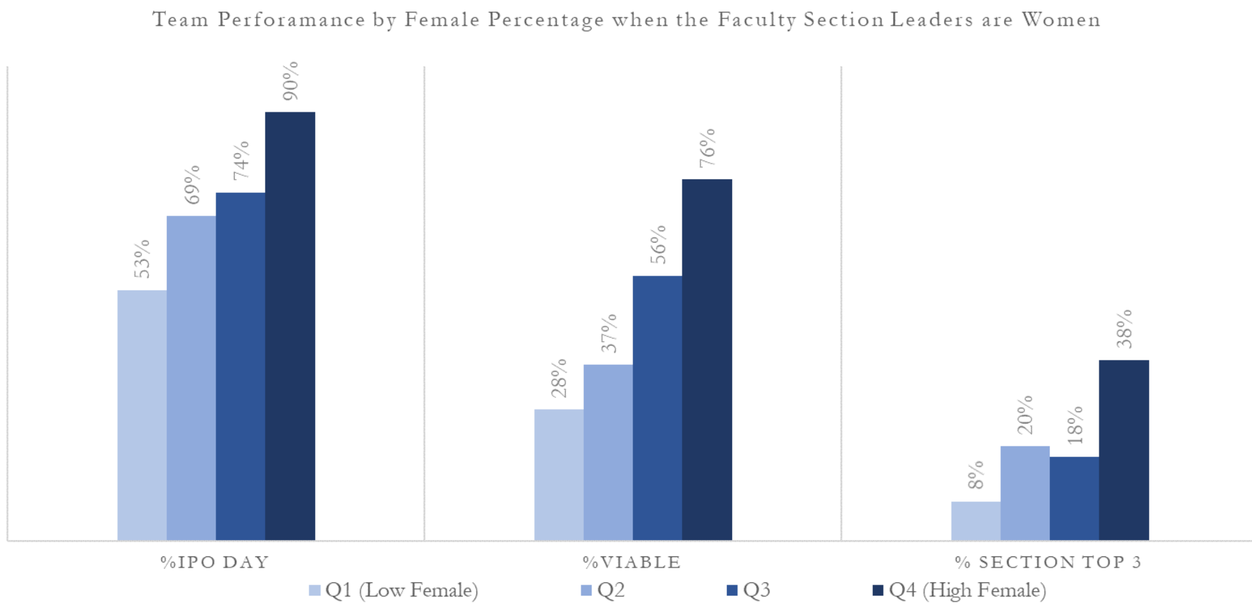
Panel A. Diversity vs. Performance with Random Team Assignment



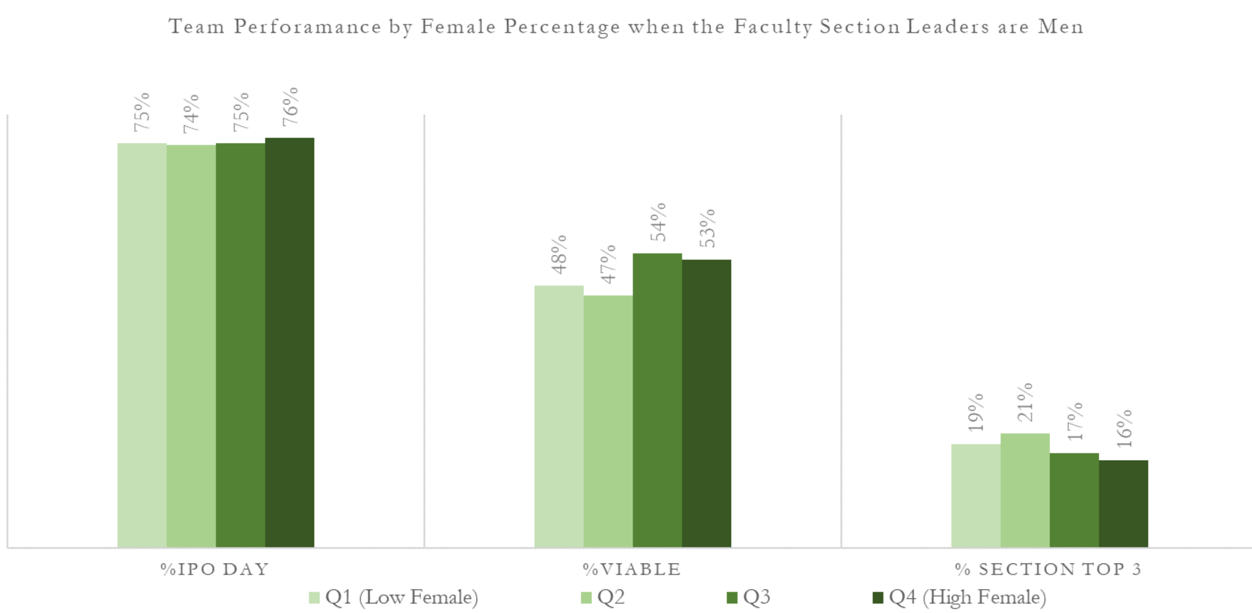
Panel B. Diversity vs. Performance with Voluntary Team Formation

Figure IV. Team Performance Conditional on Judge's Gender (2014-2016)

This figure plots the team performance conditional on judge's gender and female percent in the team. Team performance measures (Y Axis) are the percentage of teams in section top 3, viable and IPO day. Teams are sorted into four quantiles by percent female in the team. The sample includes all teams in 2014-2016. Teams in 2013 are excluded as HBS does not have judge information for that year.



Panel A. Performance Conditional on Female Section Leader and % Female Team Members



Panel B. Performance Conditional on Male Section Leader and % Female Team Members

Table I. Summary Statistics of MBA Backgrounds

The table below presents the summary statistics of the demographic and employment backgrounds of Harvard Business School MBA Class of 2013, 2014, 2015, and 2016.

	2013	**2014**	**2015**	**2016**	**Total**
# of Students	907	915	931	931	3684
Team Size	6.06	6.13	6.25	5.2	5.91
Age	28.89	29.1	29.07	29.21	29.06
% of Female	39.25%	40.44%	41.14%	41.35%	40.55%
% of White American	37.16%	39.45%	37.70%	39.53%	38.46%
% of Asian American	14.33%	11.80%	11.92%	11.82%	12.46%
% of Black	4.52%	5.68%	5.59%	5.80%	5.40%
% of Latinxs American	3.75%	4.26%	4.83%	3.65%	4.13%
% International	34.07%	34.32%	34.59%	37.06%	35.02%
Employment Background					
% Finance Background	29.66%	29.29%	33.83%	36.84%	32.44%
% Consulting	21.94%	20.55%	20.62%	25.13%	22.07%
% Technology	9.04%	9.84%	10.85%	13.96%	10.94%
% Healthcare	8.16%	7.87%	6.34%	8.92%	7.82%
Education Background					
% Ivy League	26.90%	25.03%	23.63%	22.99%	24.62%
% Top School	41.23%	37.92%	38.35%	34.26%	37.92%

Table II. Past Employment and Education Background

This table summarizes the employment and education background of HBS MBAs.

Rank	Company	Obs	Percent
1	McKinsey & Company	308	8.40%
2	Bain & Company	184	5.02%
3	Boston Consulting Group	173	4.72%
4	Goldman Sachs	166	4.53%
5	Morgan Stanley	138	3.77%
6	Google	78	2.13%
7	Credit Suisse	54	1.47%
8	J.P. Morgan	47	1.28%
9	Deloitte Consulting	45	1.23%
10	Booz & Company	44	1.20%
11	UBS Investment Bank	42	1.15%
12	Bank of America Merrill Lynch	38	1.04%
13	Bain Capital	32	0.87%
14	United States Marine Corps	29	0.79%
15	Accenture	26	0.71%
15	Citigroup	26	0.71%
15	Barclays Capital	25	0.68%
15	Oliver Wyman	25	0.68%
15	The Blackstone Group	25	0.68%
20	Deutsche Bank	24	0.65%
20	The Carlyle Group	24	0.65%
	Top 20 Total	1553	42.37%
	Sample Total	3,665	

Rank	School	Obs	Percent
1	Harvard University	286	8.17%
2	Stanford University	157	4.49%
3	University of Pennsylvania	151	4.31%
4	Yale University	124	3.54%
5	Princeton University	102	2.91%
6	Duke University	81	2.31%
7	MIT	72	2.06%
8	United States Military Academy	70	2.00%
9	Dartmouth College	67	1.91%
10	University of California	64	1.83%
11	Cornell University	63	1.80%
12	Georgetown University	60	1.71%
13	Brown University	57	1.63%
13	Columbia University	57	1.63%
15	Northwestern University	56	1.60%
16	University of Virginia	52	1.49%
17	Indian Institute of Technology	50	1.43%
18	University of Texas	45	1.29%
19	University of Michigan	38	1.09%
20	Brigham Young University	37	1.06%
	Top 20 Total	1689	48.26%
	Sample Total	3,500	

Table III. Summary Statistics on Team Performance Measures

This table reports our performance measure and the percentage of teams presented on IPO day, and ranked viable, section top 3, or class top 3.

Class Year	Freq.	IPO Day	Viable	Section Top 3	Class Top 3	Performance	SD
2013	150	78.67%	46.67%	20.00%	2.67%	0.502	0.275
2014	150	70.00%	39.33%	20.00%	2.00%	0.46	0.29
2015	150	73.33%	55.33%	20.00%	2.00%	0.512	0.287
2016	180	76.11%	52.78%	16.67%	2.22%	0.504	0.272
Total	630	74.60%	48.73%	19.05%	2.22%	0.495	0.281

Table IV. IPO Day Judge Characteristics

This table reports summary statistics on judges' gender and race/ethnicity. Each section has one section leader judge, who is a faculty member from HBS, and 3-4 other judges from the industry.

Class Year	# Judges	% Female	% Black	% Latinx	% East Asian	% South Asian	% White
Section Leader Judges							
2014	10	30%	20%	10%	0%	10%	60%
2015	10	30%	10%	0%	0%	20%	70%
2016	10	20%	20%	20%	0%	0%	60%
All Judges							
2014	49	14.29%	6.12%	4.08%	6.12%	6.12%	77.55%
2015	43	27.91%	6.98%	0.00%	9.30%	9.30%	74.42%
2016	44	34.09%	11.36%	4.55%	4.55%	6.82%	68.18%

Table V. Matching Properties of Computer-Assigned Teams for Class of 2013

This table reports the regression results of matching on race/ethnicity (gender, school, industry) ties. Each observation is a student-student pair. The dependent variable Real Match equals one if the pair is in the same team. The independent variables race/ethnicity (gender, school, industry) tie equals one if the pair has the same race/ethnicity (gender, school, industry). In addition, Both Non-US Citizens is an indicator variable equal to one if the student pairs are non-US citizens. Race/ethnicity Tie (US) is an indicator variable equal to one if the student pairs are both US citizens with the same race/ethnicity. Race/ethnicity Tie (International) is an indicator variable equal to one if the student pairs are both non-US citizens from the same region. Robust standard errors are clustered at the student level.

	Dependent Variable: Real Match					
	(1)	(2)	(3)	(4)	(5)	(6)
Subsample	Full Sample	Same Gender	Different Gender	Both US	Both Non-US	US, Non-US pairs
Gender Tie	-0.017***			-0.016***	-0.019***	-0.017***
	(0.001)			(0.002)	(0.004)	(0.002)
Both Non-US Citizens	-0.0094***	-0.010***	-0.0084**			
	(0.002)	(0.003)	(0.004)			
Race/ethnicity Tie (US)	-0.00053	0.0012	-0.0025	0.0029		
	(0.002)	(0.003)	(0.003)	(0.002)		
Race/ethnicity Tie (International)	-0.0063	-0.0058	-0.0066		-0.0069	
	(0.006)	(0.007)	(0.010)		(0.006)	
School Tie	-0.0025	0.0030	-0.0087	0.0019	0.0082	-0.011
	(0.006)	(0.008)	(0.009)	(0.007)	(0.018)	(0.011)
Industry Tie	-0.000011	-0.0013	0.0013	-0.0010	0.0092	-0.0014
	(0.002)	(0.003)	(0.003)	(0.003)	(0.006)	(0.003)
Team Member Count	0.011***	0.010***	0.011***	0.015***	0.015***	0.0057***
	(0.000)	(0.001)	(0.001)	(0.002)	(0.005)	(0.002)
Constant	0.0017	-0.013**	-0.00022	-0.032***	-0.034	0.035***
	(0.002)	(0.006)	(0.006)	(0.011)	(0.032)	(0.013)
Observations	81,368	42,140	39,228	35,228	6,764	39,376
R-squared	0.002	0.000	0.000	0.002	0.003	0.001
Year FE	YES	YES	YES	YES	YES	YES

Table VI. Matching Regression

This table reports the regression results of matching on race/ethnicity (gender, education, industry) ties. Each observation is a student-student pair. The dependent variable *Real Match* equals one if the pair is in the same team. In columns 1 and 3, the independent variables *race/ethnicity (gender, education, industry) tie* equals one if the pair has the same race/ethnicity (gender, education, industry). In columns 2 and 4, the independent variable *Endowed Demographic Match* (*Acquired Characteristics Match*) equals one if the pair has the same race/ethnicity or gender (education or industry background). Robust standard errors are clustered at the student level (i.e., one student is matched to 89 potential matches, and they are treated as one cluster).

	Dependent variable: Real Match			
	(1) Voluntarily Formed (2014-2016)	(2) Randomly Assigned (2013)	(3) Voluntarily Formed (2014-2016)	(4) Randomly Assigned (2013)
Race/ethnicity Tie	0.014***	-0.00084		
	(0.001)	(0.002)		
Gender Tie	0.013***	-0.017***		
	(0.001)	(0.001)		
School Tie	0.0085**	-0.0028		
	(0.004)	(0.006)		
Industry Tie	0.0062***	-0.00027		
	(0.001)	(0.002)		
Endowed Demographic Match			0.015***	-0.015***
			(0.001)	(0.001)
Acquired Characteristics Match			0.0066***	-0.00039
			(0.001)	(0.002)
Team Member Count	0.011***	0.011***	0.011***	0.011***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-0.019***	0.00088	-0.019***	0.00076
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	254,318	81,368	254,318	81,368
R-squared	0.003	0.001	0.002	0.001
Year FE	YES	YES	YES	YES

Table VII. Detailed Matching Regression: On Ethnicity and Gender Groups

This table reports the regression results of the probability of match on race/ethnicity ties and gender ties. Each observation is a student-student pair. The dependent variable real match equals one if the students are teammates. The independent variables are race/ethnicity, or gender ties equal one if both students share the same race/ethnicity or gender. Robust standard errors are clustered at the student level.

	Dependent variable: Real Match			
	(1)	(2)	(3)	(4)
	Voluntarily Formed (2014-2016)		Randomly Assigned (2013)	
Both White	0.012***		-0.00024	
	(0.001)		(0.002)	
Both Asian American	0.014***		0.0019	
	(0.004)		(0.005)	
Both Latinx American	0.0031		0.0041	
	(0.012)		(0.022)	
Both Black	0.013		-0.000034	
	(0.009)		(0.018)	
Both International (same region)	0.040***		-0.016***	
	(0.005)		(0.005)	
Both Male		0.012***		-0.014***
		(0.001)		(0.001)
Both Female		0.017***		-0.022***
		(0.002)		(0.001)
Team Member Count	0.011***	0.011***	0.011***	0.011***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-0.012***	-0.016***	-0.0077***	0.00064
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	254,318	254,318	81,368	81,368
R-squared	0.002	0.002	0.000	0.002
Year FE	YES	YES	YES	YES

Table VIII. Detailed Matching Regression: On Education and Industry Backgrounds

This table reports the regression results of the probability of match on education ties and industry ties. Each observation is a student-student pair. The dependent variable real match equals one if the students are teammates. The independent variables are industry or education ties equals one if both students share the same education or industry background. Robust standard errors are clustered at the student level.

	Dependent variable: Real Match			
	(1)	(2)	(3)	(4)
	Voluntarily Formed (2014-2016)		Randomly Assigned (2013)	
Both Ivy School	0.0023		0.0061	
	(0.005)		(0.009)	
Both Non-Ivy School	0.019***		-0.014*	
	(0.006)		(0.008)	
Both Finance Industry		0.0042***		-0.00084
		(0.001)		(0.003)
Both Tech Industry		0.0046		0.021**
		(0.004)		(0.010)
Both Consulting Industry		0.0043**		-0.0052
		(0.002)		(0.004)
Both Other Industries		0.022***		0.0068
		(0.004)		(0.005)
Team Member Count	0.011***	0.011***	0.011***	0.011***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-0.0096***	-0.010***	-0.0079***	-0.0082***
	(0.000)	(0.000)	(0.000)	(0.001)
Observations	254,318	254,318	81,368	81,368
R-squared	0.001	0.001	0.000	0.000
Year FE	YES	YES	YES	YES

Table IX. Match Regression: Differences by Gender

This table reports the regression results of the probability of match on race/ethnicity ties, education ties, and industry ties by gender. Each observation is a student-student pair. The dependent variable real match equals one if the students are teammates. Columns 1 and 3 report the results of the male student subsample. Columns 2 and 4 report the results of the female student subsample. Robust standard errors are clustered at the student level.

	Dependent variable: Real Match			
	(1)	(2)	(3)	(4)
	Voluntarily Formed (2014-2016)		Randomly Assigned (2013)	
	Male	Female	Male	Female
Race/ethnicity Tie	0.015***	0.011***	-0.0024	0.00088
	(0.002)	(0.002)	(0.002)	(0.003)
School Tie	0.015***	0.00018	-0.0031	-0.0025
	(0.005)	(0.005)	(0.008)	(0.009)
Industry Tie	0.0087***	0.0031*	-0.0033	0.0039
	(0.002)	(0.002)	(0.003)	(0.003)
Team Member Count	0.010***	0.011***	0.011***	0.011***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-0.011***	-0.011***	0.0063***	0.0080***
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	150,093	104,225	49,434	31,934
R-squared	0.002	0.002	0.000	0.000
Year Fixed Effect	Y	Y	Y	Y

Table X. Impact of Team Diversity on Performance

This table regresses team performance on team diversity scores. The dependent variable *Performance* is the median of the quantile of the team's outcome. The independent variables are diversity scores described in the paper. *Voluntarily Formed* is an indicator variable equal to one if the team is in 2014-2016 subsample. Robust standard error is clustered at year-section level. Statistical significance at the 1%, 5%, and 10% levels are denoted by ***, **, and *, respectively.

	Dependent variable: Performance					
	(1)	(2)	(3)	(4)	(5)	(6)
	Randomly Assigned (2013)		Organically Formed (2014-2016)		Full Sample	
Race/ethnicity Score	-0.49***	-0.45***	-0.18***	-0.17***	-0.49***	-0.46***
	(0.105)	(0.096)	(0.056)	(0.060)	(0.103)	(0.097)
Gender Score	-0.025	-0.080	-0.043	-0.064	-0.00037	-0.057
	(0.658)	(0.627)	(0.059)	(0.061)	(0.630)	(0.603)
School Score	-0.96*	-0.81	0.20	0.30	-0.97**	-0.80*
	(0.476)	(0.469)	(0.295)	(0.315)	(0.455)	(0.441)
Industry Score	-0.084	-0.072	0.12	0.11	-0.085	-0.064
	(0.165)	(0.179)	(0.082)	(0.084)	(0.156)	(0.164)
Voluntarily Formed ×Race/ethnicity Score					0.30**	0.29**
					(0.117)	(0.114)
Voluntarily Formed × Gender Score					-0.043	-0.0076
					(0.633)	(0.606)
Voluntarily Formed × School Score					1.16**	1.09**
					(0.541)	(0.528)
Voluntarily Formed × Industry Score					0.20	0.17
					(0.177)	(0.186)
Top School Ratio		0.084		0.091		0.088*
		(0.079)		(0.060)		(0.050)
Start-up Ratio		0.53		0.34**		0.37***
		(0.363)		(0.138)		(0.126)
Team Member Count	0.054	0.052	0.094***	0.085***	0.090***	0.081***
	(0.057)	(0.066)	(0.027)	(0.029)	(0.025)	(0.027)
Observations	150	150	480	480	630	630
R-squared	0.107	0.122	0.056	0.074	0.067	0.084
Year FE	N/A	N/A	YES	YES	YES	YES

Table XI. Impact of the Intersection of Gender and Race/Ethnicity on Performance

This table regresses team performance on team Ethnicity-Gender scores. The dependent variable Performance is the median of the quantile of the team's outcome. The independent variables are Ethnicity-Gender scores measure the fraction of student pairs who do not match on either gender or race/ethnicity. Voluntarily Formed is an indicator variable equal to one if the team is in 2014-2016 subsample.

	Dependent variable: Performance					
	(1)	(2)	(3)	(4)	(5)	(6)
Subsamples	Randomly Assigned (2013)		Organically Formed (2014-2016)		Full Sample	
Race/Ethnicity Score	-0.061	-0.0059	-0.17	-0.15	-0.100	-0.034
	(0.144)	(0.200)	(0.121)	(0.123)	(0.153)	(0.203)
Race/Ethnicity-Gender Score	-0.82***		0.014		-0.77***	
	(0.230)		(0.157)		(0.237)	
Race/Ethnicity-Gender Score (Male)		-0.84***		0.018		-0.79***
		(0.246)		(0.155)		(0.254)
Race/Ethnicity-Gender Score (Female)		-1.23		-0.15		-1.24*
		(0.748)		(0.213)		(0.706)
Race/Ethnicity Score * Voluntarily Formed					-0.065	-0.11
					(0.195)	(0.237)
Race/Ethnicity-Gender Score * Voluntarily Formed					0.78***	
					(0.286)	
Race/Ethnicity-Gender Score (Male) * Voluntarily Formed						0.81**
						(0.299)
Race/Ethnicity-Gender Score (Female) * Voluntarily Formed						1.09
						(0.737)
Gender Score	0.018	-0.12	-0.056	-0.056	0.081	-0.077
	(0.610)	(0.678)	(0.072)	(0.073)	(0.588)	(0.641)
Gender Score * Voluntarily Formed					-0.14	0.017
					(0.591)	(0.644)
Top School Ratio	0.20**	0.18**	0.075	0.069	0.100**	0.093*
	(0.074)	(0.075)	(0.059)	(0.058)	(0.048)	(0.047)
Start-up Ratio	0.48	0.47	0.36**	0.35**	0.36***	0.36***
	(0.317)	(0.324)	(0.139)	(0.140)	(0.126)	(0.127)
Team Member Count	0.058	0.052	0.079***	0.078**	0.075***	0.074***
	(0.070)	(0.068)	(0.028)	(0.029)	(0.026)	(0.026)
Observations	150	150	480	480	630	630
R-squared	0.141	0.145	0.067	0.070	0.083	0.085
Year FE	N/A	N/A	YES	YES	YES	YES

Table XII. Impact of Team Diversity on Performance by Subgroups

This table regresses team performance on team diversity scores. The dependent variable *Performance* is the median of the quantile of the team's outcome. The independent variables are diversity scores that measure the fraction of student pairs in each team that does not belong to the same racial or ethnic group. Robust standard error is clustered at year-section level. Statistical significance at the 1%, 5%, and 10% levels are denoted by ***, **, and *, respectively.

	Dependent Variable: Performance	
	(1) Randomly Assigned (2013)	(2) Voluntarily Formed (2014-2016)
White American Score	-0.50***	-0.19***
	(0.107)	(0.054)
Asian American Score	-1.08***	-0.013
	(0.325)	(0.230)
African American Score	-0.091	0.033
	(1.804)	(0.732)
Latinx American Score	-2.58	-1.32
	(1.930)	(0.989)
International Student Score	-1.86***	0.083
	(0.391)	(0.172)
Top School Ratio	0.11	0.072
	(0.086)	(0.049)
Startup Ratio	0.48	0.39***
	(0.374)	(0.128)
Team Member Count	0.072	0.074***
	(0.077)	(0.025)
Constant	5.97	1.38
	(4.191)	(1.365)
Observations	150	630
R-squared	0.152	0.076
Year FE	YES	YES

Table XIII. The Effect of Judge Gender on Performance

This table regresses team performance on judge's gender interacted with percent of female in the team. The dependent variable *Performance* is the median of the quantile of the team's outcome. The independent variable *Section Leader Female* is an indicator variable equals to 1 if the female judge is a section leader. *Have Female Judge* equals 1 if there is at least one female judge in the section. *Female Team Member%* is the percent of females in the team. Robust standard error is clustered at year-section level. Statistical significance at the 1%, 5%, and 10% levels are denoted by ***, **, and *, respectively.

	Dependent Variable: Performance		
	(1)	(2)	(3)
		2014-2016	
Section Leader Female * Female Team Member %	0.32***		0.33***
	(0.092)		(0.101)
Section Leader Female	-0.15***		-0.16***
	(0.043)		(0.046)
Female Team Member %	0.011	0.042	0.043
	(0.064)	(0.102)	(0.103)
Have Female Judge * Female Team Member %		0.068	-0.039
		(0.120)	(0.131)
Have Female Judge		-0.029	0.026
		(0.055)	(0.057)
Top School Ratio	0.089	0.085	0.090
	(0.059)	(0.061)	(0.060)
Start-up Ratio	0.35**	0.35**	0.35**
	(0.135)	(0.135)	(0.135)
Team Member Count	0.086***	0.089***	0.085***
	(0.029)	(0.028)	(0.029)
Constant	-0.10	-0.14	-0.12
	(0.171)	(0.167)	(0.170)
Observations	480	480	480
R-squared	0.076	0.059	0.076
Year FE	YES	YES	YES

Online Appendix

Appendix Table I. Home Country of HBS MBA Students

This table reports the top 20 home countries of HBS MBA students in our sample.

	Country	Freq.	Percent
1	USA	2,394	64.98%
2	India	172	4.67%
3	Canada	125	3.39%
4	China	76	2.06%
5	UK	59	1.60%
6	Brazil	52	1.41%
7	Australia	46	1.25%
8	France	45	1.22%
9	Germany	45	1.22%
10	Israel	33	0.90%
11	Korea	30	0.81%
12	Japan	28	0.76%
13	Mexico	28	0.76%
14	Turkey	28	0.76%
15	Argentina	27	0.73%
16	Lebanon	25	0.68%
17	Russia	25	0.68%
18	Spain	24	0.65%
19	Nigeria	23	0.62%
20	Chile	19	0.52%
	Total	3,684	100.00%

Appendix Table II. HBS Section Assignment in 2013-2016

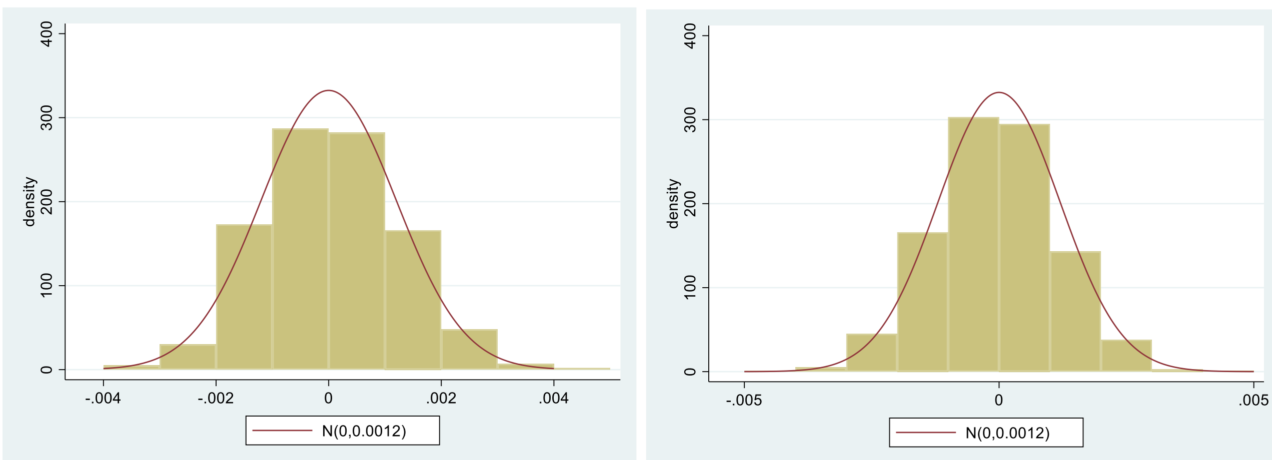
This table reports the regression results of matching on race/ethnicity (gender, education, industry) ties within sections. Each observation is a student-student pair in the section. The dependent variable Real Match equals one if the pair is in the same section. The independent variables race/ethnicity (gender, education, industry) tie equals one if the pair has the same race/ethnicity (gender, education, industry). In addition, Both Non-US Citizens is an indicator variable equal to one if the student pairs are non-US citizens. Robust standard errors are clustered at the student level.

	Dependent Variable: Real Match			
	(1)	(2)	(3)	(4)
Class Year	2013	2014	2015	2016
Gender Tie	-0.0018***	-0.0018***	-0.0018***	-0.0019***
	(0.000)	(0.000)	(0.000)	(0.000)
Both Non-US Citizens	-0.00012	0.00061	-0.00028	-0.00053
	(0.001)	(0.001)	(0.001)	(0.000)
Race/ethnicity Tie (US)	-0.0018***	-0.0021***	-0.0014**	-0.0023***
	(0.001)	(0.000)	(0.001)	(0.000)
Race/ethnicity Tie (International)	-0.0085***	-0.012***	-0.011***	-0.012***
	(0.002)	(0.002)	(0.001)	(0.001)
School Tie	-0.013***	-0.014***	-0.015***	-0.011***
	(0.002)	(0.002)	(0.001)	(0.002)
Industry Tie	-0.0031***	-0.0033***	-0.0030***	-0.0026***
	(0.001)	(0.001)	(0.000)	(0.000)
Team Member Count	0.00079***	0.00032***	0.00075***	0.00037***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-0.042***	0.044***	-0.039***	0.033***
	(0.005)	(0.005)	(0.011)	(0.011)
Observations	821,742	836,310	865,830	865,830
R-squared	0.000	0.000	0.000	0.000

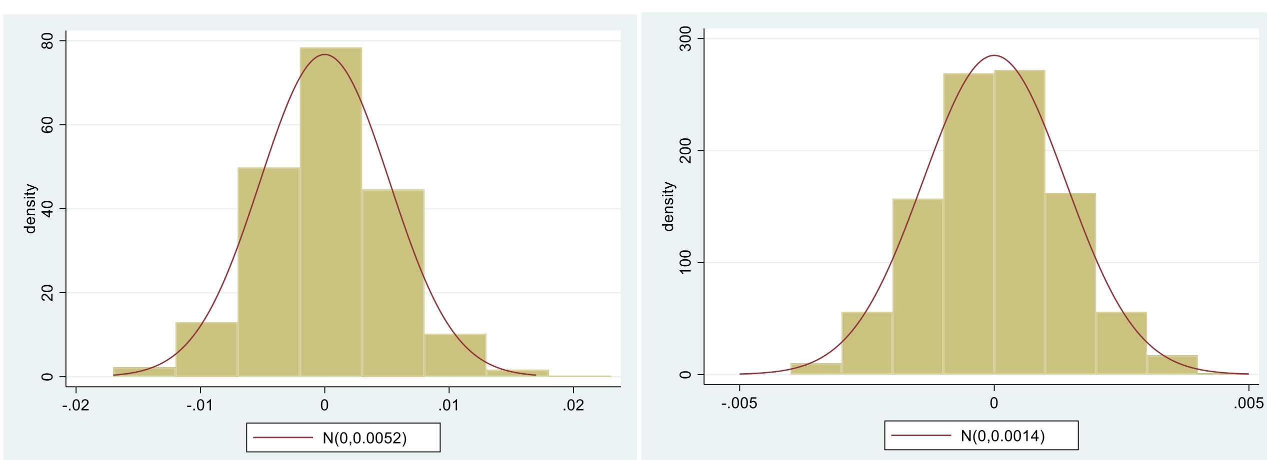
Appendix Table III. Randomization Test of Homophily

This table compares simulated matching results for 2014-2016 matching with random team assignment to actual regression coefficients. For simulated coefficients, we use 1000 iterations and report the mean and standard deviations of simulated regression coefficients. P-value is the probability of the simulated value greater than the actual matching coefficient.

Subsample: 2014-2016	Simulated Matching Coefficients		Actual Matching Coefficients		
Coefficients	Mean	SD	Value	SE	p-value
Race/Ethnicity Tie	0.0000	0.0012	0.0135	0.0011	<1%
Gender Tie	-0.0001	0.0012	0.0131	0.0011	<1%
School Tie	0.0002	0.0052	0.0085	0.0038	<5%
Industry Tie	0.0001	0.0014	0.0062	0.0012	<1%



Panel A. Distribution of Race/Ethnicity Tie Coefficients (left) and Gender Tie Coefficients (Right)



Panel B. Distribution of Race/Ethnicity Tie Coefficients (left) and Gender Tie Coefficients (Right)

Appendix Table IV. Match between International Students

This table reports the regression results of the probability of match among international students. Each observation is a student-student pair. The dependent variable real match equals to 1 if the students are teammates. The independent variables are ethnicity characteristics, equaling to 1 if both students are from the same region. Robust standard errors are clustered at the student level.

	Dependent Variable: Real Match	
	(1)	(2)
	Voluntarily Formed Teams (2014-2016)	Randomly Assigned Teams (2013)
Both European	0.025***	-0.0089
	(0.007)	(0.008)
Both South Asia	0.029***	-0.027***
	(0.008)	(0.008)
Both East Asia	0.063***	-0.040***
	(0.016)	(0.011)
Both Latin America	0.062***	-0.011
	(0.019)	(0.022)
Both Middle East	0.067***	0.043
	(0.018)	(0.048)
Both African	-0.047***	-0.058***
	(0.001)	(0.001)
Team Member Count	0.011***	0.011***
	(0.000)	(0.000)
Constant	-0.011***	-0.0075***
	(0.000)	(0.001)
Observations	254,318	81,368
R-squared	0.002	0.000
Year FE	YES	YES

Appendix Table V. Detailed Match Regression: Differences by Gender

This table reports the regression results of the probability of match on race/ethnicity ties, education ties, and industry ties by gender. Each observation is a student-student pair. The dependent variable real match equals one if the students are teammates. Columns 1 to 3 report the results of the male student subsample. Columns 4 to 6 report the results of the female student subsample. Robust standard errors are clustered at the student level.

	Dependent variable: Real Match					
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Male Subsample			Female Subsample		
Both White	0.014***			0.0091***		
	(0.002)			(0.002)		
Both Asian American	0.015**			0.014***		
	(0.007)			(0.005)		
Both Latinx American	0.016			-0.025		
	(0.016)			(0.019)		
Both Black	0.0033			0.024*		
	(0.011)			(0.014)		
Both International	0.043***			0.036***		
	(0.006)			(0.008)		
Both Ivy School		0.0089			-0.0039	
		(0.007)			(0.007)	
Both Non-Ivy School		0.025***			0.0090	
		(0.008)			(0.009)	
Both Finance Industry			0.0080***			-0.0017
			(0.002)			(0.002)
Both Tech Industry			0.0087*			-0.0022
			(0.005)			(0.006)
Both Consulting Industry			0.0026			0.0056**
			(0.003)			(0.003)
Both Other Industries			0.021***			0.023***
			(0.005)			(0.007)
Team Member Count	0.011***	0.011***	0.011***	0.011***	0.011***	0.011***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.012***	-0.0096***	-0.0100***	-0.012***	-0.0095***	0.011***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)
Observations	150,093	150,093	150,093	104,225	104,225	104,225
R-squared	0.002	0.001	0.001	0.002	0.001	0.001
Year Fixed Effect	Y	Y	Y	Y	Y	Y

Appendix Table VI. Matching Between Venture Capital Investor and Entrepreneurs

In this table, we analyze all venture capital deals in the US between 2010 and 2016, covering 5,731 venture capital investors and 11,471 deals. In the regression tables, each observation is a VC-entrepreneur pair, where each VC-deal is matched to all pseudo-deals in the same year-industry. The independent variables Same Ethnicity (Same School) are binary indicators. Robust standard errors are clustered at the VC-deal level.

Panel A. VC-Entrepreneur Match by Demographic Characteristics

	(1)
Dependent Variable	Real Match
Both Female	0.0045***
	(0.001)
Both Male	0.00098***
	(0.000)
Same Ethnicity	0.0025***
	(0.000)
Same School	0.013***
	(0.001)
Number of Pseudo Deals	-0.000***
	(0.000)
Observations	3,937,141
R-squared	0.005
Year FE	YES

Panel B. VC-Entrepreneur Match by Race / Ethnicity: Gender Differences

	Dependent variable: Real Match	
	(1)	(2)
Subsample	Male	Female
Both White	0.0017***	0.00033
	(0.000)	(0.001)
Both Jewish	0.0034***	0.0010
	(0.001)	(0.002)
Both East Asian	0.022***	0.016***
	(0.002)	(0.004)
Both Indian	0.015***	0.011***
	(0.001)	(0.003)
Both Hispanic	0.013***	0.030
	(0.004)	(0.023)
Both Black	0.14	
	(0.098)	
Observations	3,609,876	291,338
R-squared	0.006	0.006
Year FE	YES	YES

Appendix Table VII. Impact of Team Diversity on Performance with Additional Controls

This table regresses team performance on team diversity scores with various controls. The dependent variable *Performance* is the median of the quantile of the team's outcome. The independent variables are diversity scores described in the paper. Robust standard errors are clustered at year-section level. Each coefficient's standard error appears directly below the coefficient estimate. Statistical significance at the 1%, 5%, and 10% levels are denoted by ***, **, and *, respectively.

	Dependent Variable: Performance				
	(1)	(2)	(3)	(4)	(5)
Subsample	Randomly Assigned (2013)				
Race/ethnicity Score	-0.68**	-0.51**	-0.44***	-0.45***	-0.46***
	(0.257)	(0.158)	(0.105)	(0.097)	(0.103)
Gender Score	-0.13	-0.11	-0.099	-0.035	-0.067
	(0.639)	(0.667)	(0.587)	(0.631)	(0.626)
School Score	-0.82	-0.84	-0.83*	-0.81	-0.80
	(0.467)	(0.489)	(0.443)	(0.464)	(0.457)
Industry Score	-0.019	-0.056	-0.046	-0.16	-0.073
	(0.196)	(0.169)	(0.205)	(0.158)	(0.179)
Top School Ratio	0.072	0.088	0.084	0.093	0.084
	(0.081)	(0.081)	(0.079)	(0.081)	(0.078)
Start-up Ratio	0.49	0.57	0.53	0.54	0.51
	(0.380)	(0.389)	(0.360)	(0.374)	(0.368)
Team Member Count	0.056	0.052	0.050	0.052	0.053
	(0.068)	(0.066)	(0.063)	(0.065)	(0.065)
White%	-0.22				
	(0.191)				
English Speaker%		-0.14			
		(0.191)			
Consulting%			0.063		
			(0.194)		
Finance%				-0.11	
				(0.176)	
Technology%					0.095
					(0.082)
Constant	1.62	1.56	1.38	1.46	1.36
	(1.064)	(1.065)	(0.917)	(0.833)	(0.898)
Observations	150	150	150	150	150
R-squared	0.130	0.126	0.123	0.125	0.124

Appendix Table VIII. Correlation between Industry Experience and Demographic Characteristics

This table regresses past industry employment on student gender, race, and ethnicity. Each observation is a student. The analysis includes all students from 2013 to 2016.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Finance	Technology	Consulting	Healthcare	Retail	Top School	Start-up Exp
Female	-0.020	-0.0043	0.094***	5.3e-06	0.0096	0.10***	0.0054
	(0.016)	(0.011)	(0.014)	(0.009)	(0.006)	(0.016)	(0.007)
Asian American	0.046*	0.061***	-0.028	0.032*	0.00017	0.14***	0.0056
	(0.025)	(0.019)	(0.020)	(0.017)	(0.010)	(0.026)	(0.012)
African American	0.052	-0.028	0.010	-0.041**	0.027	0.026	-0.022*
	(0.036)	(0.019)	(0.030)	(0.018)	(0.017)	(0.038)	(0.012)
Hispanic American	-0.021	-0.0051	0.011	-0.017	0.014	-0.047	-0.0067
	(0.039)	(0.025)	(0.032)	(0.023)	(0.018)	(0.041)	(0.017)
International	-0.035**	0.017	0.13***	-0.056***	-0.00052	-0.29***	-0.017**
	(0.017)	(0.012)	(0.016)	(0.009)	(0.007)	(0.016)	(0.007)
Constant	0.31***	0.079***	0.14***	0.099***	0.029***	0.45***	0.031***
	(0.018)	(0.011)	(0.016)	(0.011)	(0.007)	(0.019)	(0.007)
Observations	3,684	3,684	3,684	3,684	3,684	3,684	3,684
R-squared	0.008	0.008	0.039	0.015	0.002	0.121	0.007
Year FE	YES	YES	YES	YES	YES	YES	YES

Appendix Table IX. Impact of Team Diversity on Performance: Binary Outcomes

This table reports logit regression results on the effect of Race/ethnicity-Gender Score. Robust standard error is clustered at year-section level. The dependent variables IPO day/Viable/Section Top 3/Class year top 3 are indicator variables equals 1 if the team presented on IPO day/the project is deemed viable by judges/the team is section top 3/the team is class year top 3. The independent variables are diversity scores described in the paper. Robust standard error is clustered at year-section level. Each coefficient's standard error appears directly below the coefficient estimate. Statistical significance at the 1%, 5%, and 10% levels are denoted by ***, **, and *, respectively.

Panel A. Random Assignment (2013)	(1)	(2)	(3)	(4)
VARIABLES	IPO Day	Viable	Section Top 3	Class Top 3
Race/Ethnicity-Gender Score	-10.45***	-6.955***	-4.002**	-6.209***
	(3.105)	(1.998)	(1.575)	(1.465)
Top School Ratio	1.772**	1.289**	1.267	0.559
	(0.830)	(0.533)	(0.982)	(1.568)
Start-up Ratio	2.819	3.613**	3.428	2.246
	(2.964)	(1.766)	(3.318)	(8.201)
Team Member Count	1.103	0.397	-0.153	-0.867
	(0.738)	(0.626)	(0.645)	(0.591)
Constant	3.374	2.984	2.367	6.600
	(5.372)	(4.485)	(3.266)	(4.045)
Observations	150	150	150	150
Panel B. Voluntary Formation (2014-2016)	(5)	(6)	(7)	(8)
VARIABLES	IPO Day	Viable	Section Top3	Class Top 3
Race/Ethnicity-Gender Score	-1.927**	-1.066**	-1.291**	-2.319
	(0.803)	(0.472)	(0.620)	(1.438)
Top School Ratio	0.784	0.677*	0.102	-0.288
	(0.510)	(0.352)	(0.546)	(1.052)
Start-up Ratio	2.257*	2.772***	2.966***	
	(1.286)	(1.023)	(1.123)	
Team Member Count	0.474*	0.739***	0.304	0.766
	(0.257)	(0.250)	(0.245)	(0.673)
Constant	-0.781	-4.424***	-2.323	-6.511
	(1.501)	(1.481)	(1.698)	(4.502)
Fixed Effects	Year	Year	Year	Year
Observations	480	480	480	375

Appendix Table X. Alternative Measure of Race/Ethnicity Diversity

This table regresses team performance on team diversity scores. The dependent variable *Performance* is the median of the quantile of the team's outcome. The independent variables *Race/ethnicity Percentile* is the median percentile of the simulated distribution of the Race/ethnicity score. We obtain the simulated distribution by randomly assigning students to teams in each class year and computing the Race/ethnicity score in each iteration. We iterate this process for 1000 times. *Ethnicity-Gender Percentile* is constructed analogously. Robust standard error is clustered at year-section level. Each coefficient's standard error appears directly below the coefficient estimate. The Statistical significance at the 1%, 5%, and 10% levels are denoted by ***, **, and *, respectively.

	Dependent Variable: Performance			
	(1)	(2)	(3)	(4)
Subsample	2013	2014-2016	Full Sample	Full Sample
Race/ethnicity Percentile	-0.36***	-0.16***	-0.34***	-0.14
	(0.053)	(0.044)	(0.058)	(0.115)
Race/ethnicity Percentile * Voluntary Formation			0.19**	-0.076
			(0.073)	(0.134)
Ethnicity-Gender Percentile				-0.27**
				(0.124)
Ethnicity-Gender Percentile * Voluntary Formation				0.34**
				(0.145)
Gender Score	0.040	-0.049	-0.044	-0.070
	(0.673)	(0.059)	(0.059)	(0.063)
School Score	-1.06*	0.19	-0.085	-0.040
	(0.485)	(0.291)	(0.259)	(0.254)
Industry Score	-0.13	0.13	0.082	0.081
	(0.142)	(0.082)	(0.073)	(0.074)
Team Member Count	0.055	0.089***	0.082***	0.083***
	(0.052)	(0.028)	(0.025)	(0.025)
Constant	1.48*	-0.28	0.22	0.22
	(0.693)	(0.309)	(0.284)	(0.279)
Observations	150	480	630	630
R-squared	0.144	0.064	0.073	0.081
Year FE	N/A	YES	YES	YES