Diversity and Performance in Entrepreneurial Teams [*]

Sophie Calder-Wang [†] Paul A. Gompers[‡] Kanyuan (Kevin) Huang[§]

March 10, 2024

Abstract

We study how diversity affects the performance of entrepreneurial teams. In a business course to build start-ups, we find that homophily among demographic and personal characteristics plays a significant role in team formation. Next, leveraging a cohort of students where team memberships are randomly assigned, we find racial and ethnic diversity significantly degrades performance. Yet, such negative performance effect becomes mitigated when teams are formed voluntarily, suggesting an important role played by selection on unobservables. Our findings are consistent with a model where the costs of cross-racial/ethnic collaboration outweighs the complementarity benefits of demographic diversity. These findings also suggest that policy interventions to improve diversity should consider the process by which teams are formed to avoid compromising on output.

Keywords: racial and ethnic diversity, team performance, entrepreneurship, DEI, homophily

JEL Codes: J15, J16, L26, M12, G24

[*]We are grateful to Kari Limmer at Harvard Business School MBA Student and Academic Services for providing us with extensive course information. This paper benefited from conversations with seminar and conference participants at Harvard Business School, Stanford Graduate School of Business, Carnegie Mellon Tepper School of Business, University of Utah David Eccles School of Business, Tilburg University, NBER Summer Institute (Entrepreneurship), PBCSF Tsinghua University, AFA, and Workshop on Entrepreneurial Finance and Innovation (WEFI). Patrick Sweeney, Will Levinson, and Ryan Hastings provided excellent research assistance. Support for this research was provided by the Division of Research at HBS. All errors are our own.

[†]The Wharton School, University of Pennsylvania *sophiecw@wharton.upenn.edu*

[‡]Harvard Business School and NBER *pgompers@hbs.edu*

[§]Chinese University of Hong Kong, Shenzhen *huangkanyuan@cuhk.edu.cn*

1 Introduction

Improving diversity, equity, and inclusion has increasingly become a key objective for many corporations, affecting hiring and promotion decisions at all levels in the workplace. A growing number of regulations have explicitly mandated diversity on corporate boards as a government policy, such as the gender quota in Norway (Ahern and Dittmar, 2012; Matsa and Miller, 2013; Bertrand, Black, Jensen, and Lleras-Muney, 2019) and, more recently, the diversity quota in California for under-represented minorities (California Assembly Bill 979). The SEC has also approved Nasdaq's request to mandate diversity disclosures.[1] Yet, the economic implication of diversity on firm performance is difficult to estimate empirically due to its inherent endogeneity.

In this paper, we focus on the issue of diversity in the entrepreneurial setting. Despite being the backbone of innovation and economic growth (Gornall and Strebulaev, 2021), the entrepreneurial ecosystem suffers a striking lack of diversity among start-up founders. Women make up less than 15% of start-up founders; Hispanics and Blacks make up fewer than 8% and 1% of all venture-capital-backed founders, respectively (Calder-Wang, Gompers, Huang, and Levinson, 2023). Why are entrepreneurial teams so homogeneous? Does team diversity lead to better entrepreneurial outcomes? How would policy intervention, such as mandated gender and racial quotas or funding preferences for under-represented groups, affect performance?

To make progress towards these questions, we exploit a unique series of quasi-random variations in a required first-year business course to build start-up companies. The team-based course was taken by over 3,000 Master of Business Administration (MBA) students from the Classes of 2013 through 2016 at Harvard Business School (HBS). Notably, for the Class of 2013, students were randomly assigned into teams by a computer algorithm, while for the Classes of 2014 to 2016, students were allowed to choose their own teammates. After working closely with their teammates throughout their first year spring semester, they presented their proposed start-up companies at the IPO Day and were evaluated by a panel of external judges comprised of venture capital (VC) investors and VC-backed entrepreneurs.

[1] As of August 6, 2021, Nasdaq's Board Diversity Rule requires companies listed on its exchange to publicly disclose board-level diversity statistics; and have or explain why they do not have at least two diverse directors. (See Securities Exchange Act Release No. 34-92590).

We start with a stylized model on how diversity may impact performance. Motivated by the theory literature (Lazear, 1999; Prat, 2002; Alesina and Ferrara, 2005), diversity may improve performance because of higher team quality (e.g., through complementarity in skills, knowledge, or perspectives), especially for complex, open-ended tasks, but diversity may hinder performance because the cost of exerting effective effort is higher (e.g., through challenges in communication or coordination). Because team diversity makes the cost of effort to be higher, the main testable prediction of the model is that the relative strength of the quality channel versus the cost channel determines whether diversity ultimately leads to better or worse performance. In addition, the model also predicts that if the cost channel dominates, then voluntarily formed teams will be more demographically homogeneous and more likely to match on other unobservable characteristics.

We then proceed with our empirical analysis. To begin, we provide strong evidence that homophily, namely, the phenomenon that "birds of a feather flock together" (McPherson, Smith-Lovin, and Cook, 2001), plays a significant role in team formation for the 2014-2016 cohorts, where students were free to select their teammates. We find that team formation is heavily driven by homophily among a wide set of demographic characteristics and personal backgrounds. Relative to a benchmark of random encounters, individuals are 25% more likely to form teams with those of the same race/ethnicity or gender. School ties and shared work experience increase the probability of matching by 17% and 11%, respectively. Consequently, teams formed voluntarily tend to be homogeneous in these dimensions.

Next, we estimate the relationship between entrepreneurial team diversity and performance under two different team formation mechanisms: random assignment and voluntary formation. By exploiting the variations in racial/ethnic diversity for the 2013 cohort when a computer algorithm was used to assign students to teams, we find that racially/ethnically diverse teams perform significantly worse than homogeneous teams, where a one-standard-deviation increase in team diversity leads to a 9 percentage points decline in performance percentiles, or about 18% relative to average performance. Because team assignment was entirely dictated by a computer algorithm for the 2013 cohort, the identity of one's teammates is independent of unobservable preferences or characteristics, providing us with a credible setting to causally estimate the impact of diversity on performance.

Given that the randomization takes place at the student level, rather than at the student characteristics level, in principle, we cannot directly test for mechanisms. Still, to shed some light on plausible mechanisms, we find that the negative performance of racial/ethnic diversity remain unchanged after controlling for a variety of plausible diversity "correlates", such as industry background, start-up experience, academic strength, and specific gender or race/ethnicities, effectively ruling them out as the explanation. Moreover, when we decompose the racial/ethnic diversity measure into various sub-components, we find a negative relationship between performance and mismatches in all racial or ethnic subgroups across the board, providing some suggestive evidence that having to work with someone demographically different may have posed as a meaningful barrier to effective collaboration.

Interestingly, although we find that racial and ethnic diversity in randomly assigned teams reduces performance, diverse teams formed endogenously through voluntary matching for the 2014-2016 student cohorts do not suffer nearly as much performance degradation. The negative performance gradient of racial and ethnic diversity in the 2013 cohort is alleviated by approximately 60% in the 2014-2016 cohorts when teams were formed voluntarily, suggesting endogenous team formation mitigates the detrimental effects of such "forced" diversity. To investigate why voluntary team formation alleviates the negative performance impact of diversity, we provide some suggestive evidence that, conditional on demographic characteristics, student pairs are more likely to choose to match on "unobservable" characteristics, such as shared career or personal interests, which are not typically easily observed by the econometrician or policy makers.

Tying back to our model, all of our main empirical results favor a model of team production where any potential quality channel of diversity is overwhelmed by the cost channel. We find a robust negative relationship between performance and diversity among randomly assigned teams. We also find clear evidence that teams become more homogeneous in demographic and personal characteristics when they are formed voluntarily.

While we hesitate to over-extrapolate from our empirical setting, to shed some light on external relevance, we also analyze a dataset of all VC-backed start-up founders in the US from 1990 onward. We first note that the majority of such start-ups are founded by more than one founder. Consistent with our main findings, real co-founders are also much more likely to have shared characteristics in gender, race, and ethnicity. Moreover, racial/ethnic diversity among the founding team has a negative, but statistically insignificant, correlation with company successes, also broadly consistent with the results in our empirical setting when teams are formed voluntarily.

Our paper contributes to several strands of literature. First, we contribute to the empirical research on the causal impact of diversity on team performance, especially in the entrepreneurial context. Theoretical work on the performance implication of team diversity focuses on the trade-off between knowledge complementarity and communication costs (Lazear, 1999; Prat, 2002; Alesina and Ferrara, 2005). Existing field experiments find a negative impact of diversity on worker productivity when performing a set of well-defined tasks (Hjort, 2014; Marx, Pons, and Suri, 2021; Aman-Rana, Minaudier, Alvarez Pereira, and Chaudry, 2021).[2] Our work instead focuses on the diversity implication of performing the considerably more complex task of starting an entrepreneurial venture, where, theoretically, the benefit of team diversity through knowledge complementarity could be greater. Yet, we still find that racial and ethnic diversity in teams leads to performance declines even for such complex tasks. Notably, while our findings appear to differ from Hoogendoorn, Oosterbeek, and van Praag (2013) where they find greater gender diversity leads to better business performance through a field experiment with Dutch undergraduates,[3] our results are broadly consistent with their companion paper Hoogendoorn and van Praag (2012), which also provides preliminary evidence that mixing students of Dutch descent and foreign descent has led to worse outcomes.

Moreover, our paper is unique in the literature where we observe outcomes in both voluntary and randomized assignment mechanisms in the same setting to study the performance impact of diversity. By contrast, existing empirical methods either perform observational studies of endogenously formed teams (van Knippenberg and Schippers, 2007) or conduct field experiments to perturb team diversity (Bandiera, Rasul, and Barankay, 2013; Hoogendoorn, Oosterbeek, and van Praag, 2013). Our finding highlights a material improvement in the performance-diversity gradient when the team formation mechanism is changed from exogenous assignment to voluntary formation, suggesting the large role that selection on unobservables could play in affecting performance.

[2] Hjort (2014) and Marx, Pons, and Suri (2021) study the effect of diversity on labor productivity on flower distribution and voter registration in Kenya. Aman-Rana et al. (2021) examine worker's productivity in a lab experiment of taking a half-hour multiple choice question test.

[3] Besides differences in the setting, i.e., Dutch undergrad vs. HBS MBAs, we suspect that the differences in the results may also arise from differences in ethnicity vs. gender dynamics, especially in terms of the relative trade-off between skill complementarity and cost of communication.

To our knowledge, our stylized model provides the first formal framework of team production that allows for varying degrees of selection on unobservables, providing a tractable approach in taking the existing theoretical literature to the data. Conceptually, a key takeaway from our model is that voluntary formation leads to greater selection on unobservables, which weakens the correlation between matches on observables and matches on unobservables, thus leading to a weaker relationship between performance and observed demographic diversity.

The differences in selection on unobservables also provide a plausible explanation for bridging the gap between the results found in the experimental vs. non-experimental literature. While tightly controlled field experiments tend to find a negative effect of diversity on performance, the literature has also documented notable positive relationship between diversity and performance in a variety of empirical finance settings, ranging from hedge funds (Lu, Naik, and Teo, 2024), mutual funds (Evans, Prado, Rizzo, and Zambrana, 2019; Cohen, Frazzini, and Malloy, 2008), and venture capital (Gompers, Mukharlyamov, and Xuan, 2016), where such collaboration are all formed voluntarily.

Our paper also contributes to the smaller but growing body of work investigating the race and gender gap in entrepreneurship (Ewens, 2022; Gompers and Wang, 2017). Recent studies on this topic have focused on the funding gap for individual founders (Hebert, 2020; Fairlie, Robb, and Robinson, 2022; Cook, Marx, and Yimfor, 2022), but given that over 60% of venture-capital entrepreneurial start-ups are founded by teams, the nature of team production has important implications for explaining the cause and the effect of the gap. In addition, while the literature has long recognized the importance of social networks on entrepreneurial success (Gompers, Lerner, and Scharfstein, 2005; Nanda and Sørensen, 2010; Hochberg, Ljungqvist, and Lu, 2007, 2010), relatively little is known on how entrepreneurial teams are formed. Specifically, our experimental setting allows us to quantify the significant role of homophily along demographic characteristics and personal backgrounds in the entrepreneurial team formation process.[4] Given that the pre-existing entrepreneurial ecosystem is skewed against certain demographic groups, our work lends support towards the notion that homophilous preferences can lead to persistent racial and gender gap in entrepreneurship.

[4]The existence of homophily is widely documented in various settings, such as close friendships (Marsden, 1987, 1988; Currarini, Jackson, and Pin, 2009) and professional relationships (Stolper and Walter, 2018; Kleinbaum, Stuart, and Tushman, 2013; Ruef, Aldrich, and Carter, 2003; Reagans, 2011). However, methodologically, because we can observe one's entire choice set, our analysis provides a major improvement to observational studies, which cannot differentiate the impact of homophilous preference from pre-existing personal networks.

Lastly, our paper has policy relevance when considering the effectiveness of diversity interventions in entrepreneurial finance. There has been a proliferation of efforts to improve diversity in entrepreneurial financing. For example, a recent survey found that "54% of VCs say that investing in companies with women and/or multicultural founders is a priority to their LPs." [5] However, existing studies on the effects of mandated diversity interventions have primarily focused on the impact of gender quotas on corporate boards (Ahern and Dittmar, 2012; Matsa and Miller, 2013; Bernile, Bhagwat, and Yonker, 2018; Bertrand et al., 2019). Little is known about the performance implications of such LP pressure to invest diversely. While we cannot naively extrapolate the magnitudes in our paper, we believe that the general insight is that the performance implication of diversity will depend on how such diversity policy is implemented, where we caution against overly restrictive policies which could lead to productivity losses. Moreover, policy interventions that directly target a reduction of the costs associated with working in a diverse team will likely have the positive effect of allowing any possible complementarity benefit to come through.

The remainder of the paper is organized as follows. Section 2 describes the empirical setting. Section 3 describes a stylized model that generates testable predictions on the performance impact of diversity for both exogenously assigned and endogenously formed teams. Section 4 describes the data. In section 5, we present results on homophily in team formation. In section 6, we present our main results on the performance implication of team diversity during random assignment and voluntary formation respectively. Section 7 considers the external relevance of our results. Section 8 concludes.

2 Background and Setting

Each year, approximately 900 MBA students matriculate at the Harvard Business School. They are divided into 10 sections of approximately 90 students per section.[6] All students take the same set of required courses with their same section-mates throughout their entire first year.

For MBA class years 2013 to 2016, one of the required first-year courses was called Field Immersion Experience for Leadership Development (FIELD). In the third module of the course (i.e., FIELD 3),

[5] `https://www.morganstanley.com/ideas/vc-funding-diverse-startups-2022-survey` (Accessed April 17, 2023)

[6] They are labeled as Section A, Section B, . . . , Section J. Students are randomly assigned into different sections conditioning on observable characteristics so that each section is representative of the overall student population. Past research such as Shue (2013) and Lerner and Malmendier (2013) have shown that students typically form strong connections with their section-mates after spending the entire first year together for all their classes.

which takes place in the spring semester of the first year, students were required to work in small teams to develop and launch a start-up business. According to the distributed teaching notes, FIELD 3 was designed to allow MBA students to "hone their collaborative skills while experiencing the challenge and excitement of being an entrepreneur." FIELD 3 was one of the five classes that MBA students took during the semester.

The general structure of the course was as follows:

1. At the beginning of the spring semester, students formed a team of 5 to 7 with members all from the same section. Each team was endowed with seed funding of $3,500 to $5,000.[7]

2. Throughout the semester, these student teams worked together to develop and build their start-up businesses. They were 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.

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

4. Finally, at the end of the semester on "IPO Day," the surviving teams proceeded to present their projects to a panel of external judges from the academic, corporate, entrepreneurship, and venture capital industries. These judges ranked the teams by determining whether they had demonstrated product demand, whether the business was "viable" (defined as "positive cash flow in a five-year period"), and their ability to "create the most value" (defined as the "the value of their stock holdings").

Importantly, FIELD 3 was taught to all sections of a cohort in parallel, and it was also taught to all four cohorts of MBA students using the same course structure with one important exception: Namely, for the Class of 2013, team membership was randomly assigned by a computer algorithm developed by the HBS administration. In contrast, for the Class of 2014, 2015, and 2016, teams were formed voluntarily by students, where students freely chose their teammates within the same section.

[7] In 2013 and 2014, the initial funding was $5,000. In 2015 and 2016, the initial fund was $3,500, but students are allowed to request additional funding later on during the semester. The maximum funding should not exceed $6,500.

Therefore, for the Class of 2013, because team memberships were assigned, from a student's perspective, the identity of their teammates is exogenous, free of any selection on either observable or unobservable characteristics. In addition, whether a student was subject to random assignment is also likely exogenous to student characteristics, because it would be unlikely for a student to change their matriculation decision based on one single feature of a class, out of over twenty classes they would be taking throughout the MBA program, along with many other experiences they would have at HBS. As a result, we believe there is little scope for selection into treatment, allowing us to directly compare estimates across different assignment mechanisms.

3 Stylized Model

In this section, we present a stylized model of the performance of entrepreneurial teams, creating a conceptual framework for how team performance may be affected by diversity in both exogenously assigned and endogenously formed teams. Naturally, the model cannot exhaustively capture all aspects of the team production, but the purpose of the model is to make it more transparent how to map our empirical findings into a conceptual framework.

We will first set up a baseline model when all group characteristics are perfectly observed. We then build on the model to allow for the case when the researcher can only observe a subset of characteristics, allowing us to generate testable predictions in the data.

3.1 Baseline Model

We assume that students seek to maximize utility, which is a function of the performance of the Field 3 project less the cost of effort:

$$U(g, e) = F(q(g), e) - C(e, d(g)). \tag{3.1}$$

$F(q, e)$ models team performance as driven by the quality of the team q and the amount of effort exerted e. Moreover, the quality of the team $q(\cdot)$ may be driven by the characteristics of its group members g, which can be thought of as a vector capturing all team members' characteristics. Intuitively, the quality of the team may be driven by individual student quality, such as their past start-up experience or whether they are academically strong students, but it could also be driven by the interaction of student characteristics in the team, such as the complementarity of their skills. Naturally, performance is increasing in both quality and effort.

$C(e, d(g))$ represents the cost of exerting effort e for a given group g, driven by the amount of effort e and the extent of "team discord", which is also driven by the characteristics of its group members g. Intuitively, one may think that coming from different backgrounds may contribute to "team discord" because of higher communication costs, more collaboration challenges, or fewer shared interests, etc.

Importantly, d is a modeling device to capture the notion that the cost of effort is increasing with the degree of discord d

$$\frac{\partial^2 C}{\partial d\, \partial e} > 0. \tag{3.2}$$

The positive cross-partial means that it is more costly to generate effective effort when the team is in "discord."

To simplify exposition, we let g denote a scalar parameter that enters positively into the cost of effort $d(g)$. Intuitively, one could think of g as a measure of group diversity that would drive up the cost of effort.

Lemma 3.1. *(Optimal effort is decreasing in diversity.) Under mild regularity conditions, for a given team composition g, its optimal level of effort e^* is decreasing in group diversity*

$$\frac{\mathrm{d}e^*}{\mathrm{d}g} < 0. \tag{3.3}$$

See the appendix for the proof. The choice of optimal effort has to trade off the marginal gains from increased performance with the marginal cost of exerting more effort. Hence, the fact that the cost of effort is higher for more diverse teams implies that they choose to exert less effort.

Because neither effort nor utility is directly observed, we focus our predictions on what we do observe, namely, performance F.

Proposition 3.2. *(Performance Predictions) For a given distribution of team composition, the sign of how performance F relates to group diversity g is determined by the relative strength of the quality* *channel of diversity and the effort channel of diversity as follows*

$$\frac{\mathrm{d}}{\mathrm{d}g}F^*(q(g)) = \frac{\mathrm{d}}{\mathrm{d}g}F\left(q(g), e^*(g)\right) \tag{3.4}$$

$$= \underbrace{\frac{\partial F}{\partial q}\frac{\mathrm{d}q}{\mathrm{d}g}}_{\textit{The Quality Channel}} + \underbrace{\frac{\partial F}{\partial e}\frac{\mathrm{d}e^*}{\mathrm{d}g}}_{\textit{The Effort Channel}} . \tag{3.5}$$

The result follows directly from Lemma 3.1. The key intuition is that because the optimal effort is lower in a more diverse team, the effort channel stipulates that diversity has a *negative* impact on performance, unless the quality channel (i.e., skill complementarity from a diverse team) is large enough to overcome it.[8]

Therefore, if the effort channel of diversity dominates the quality channel, namely, $-\frac{\partial F}{\partial e}\frac{\mathrm{d}e^*}{\mathrm{d}g} > \frac{\partial F}{\partial q}\frac{\mathrm{d}q}{\mathrm{d}g}$, then, performance is decreasing in diversity, namely, $\frac{\mathrm{d}}{\mathrm{d}g}F^*(q(g)) < 0$. In other words, if we find a negative relationship between performance and diversity, then it rejects the hypothesis that the quality channel of diversity is stronger than the effort channel. Conversely, if we find a positive relationship between performance and diversity, then it rejects the hypothesis that the effort channel of diversity is stronger than the quality channel.

3.2 Predictions Based on Observable Characteristics

However, the prediction derived in the previous section is not directly testable because it relies on us being able to observe the true measure of diversity g that enters into quality $q(g)$ and the cost of effort $d(g)$. To the extent that we only observe a subset (albeit still extensive) of all student characteristics, we face both a measurement problem and an endogeneity problem, which requires us to enrich our model to accommodate these econometric issues.

To address the measurement problem that the researcher may only observe a subset of relevant student characteristics, we let $g = g(g^o, g^u)$, where g^o represent the observable student characteristics and g^u represent student characteristics that are payoff relevant but unobservable to researchers. As

[8] In theory, in addition to skill complementarity, the quality channel could also be driven by a positive correlation between individual student quality and team diversity. However, empirically, we do not find any evidence of it in our setting. There is no statistically significant relationship between individual student quality and team diversity, as shown in Appendix Table B.5. Hence, for ease of exposition, we focus our discussion on the quality channel driven by complementarity.

such, we rewrite the main model of utility maximization as

$$U(g^o, g^u, e) = F(q(g^o, g^u), e) - C(e, d(g^o, g^u)). \tag{3.6}$$

To address the endogeneity problem where students may be choosing to match on various characteristics (both observable and unobservable to the researcher) when they are free to choose their teammates, we model this parsimoniously by allowing the distribution of g to differ depending on the team formation mechanism:

$$\textit{Exogenous (randomized team assignment): } (g^o, g^u) \sim \mathcal{F}^X \tag{3.7}$$

$$\textit{Endogenous (voluntary team formation): } (g^o, g^u) \sim \mathcal{F}^N. \tag{3.8}$$

Again, to simplify exposition, we let g^o and g^u be scalars that enter positively into the cost of effort $d(g^o, g^u)$. It also enters into the quality of the team $q(g^o, g^u)$.

Lemma 3.3. *(Optimal effort is decreasing in observed and unobserved diversity.) Under mild regularity conditions, for a given team composition (g^o, g^u), its optimal level of effort e^* is decreasing in both observed and unobserved dimensions of team diversity.*

$$\frac{\partial e^*(g^o, g^u)}{\partial g^o} < 0, \quad \frac{\partial e^*(g^o, g^u)}{\partial g^u} < 0. \tag{3.9}$$

The proof is a direct extension of the proof for Lemma 3.1. Next, we derive two useful properties on how g^o and g^u relate to each other.

Lemma 3.4. *(Positive correlation between observed and unobserved diversity.) For g^o and g^u that both positively enter into the cost of effort $d(g^o, g^u)$, they are positively correlated when teams are exogenously assigned.*

$$corr(g^u, g^o)_{g \sim \mathcal{F}^X} > 0. \tag{3.10}$$

See appendix for proof. When the underlying observed and unobserved characteristics are positively correlated in the underlying population, the diversity measures also become positively correlated: matches in the observed characteristics between two individuals also imply likely matches in the unobserved characteristics.[9]

Lemma 3.5. *(Selection weakens the correlation between observed and unobserved diversity) The correlation between observed and unobserved measures of diversity becomes weakened when teams are endogenously formed:*

$$corr(g^u, g^o)_{g \sim \mathcal{F}^X} > corr(g^u, g^o)_{g \sim \mathcal{F}^N}. \tag{3.11}$$

See appendix for proof. For a team of given diversity g^o in observed characteristics, to reduce the cost of effort, they will choose to be better matched in unobserved characteristics than from a random draw from the underlying distribution. Intuitively, this means that seemingly diverse teams can have good matches on unobservables, such as shared career interests. Meanwhile, to improve potential complementarity, students can also select on unobserved characteristics to incorporate more complementary skills. Intuitively, this means that seemingly homogeneous teams can have more complementary skills, such as different industry backgrounds. In either case, allowing for selection weakens the unconditional positive correlation between g^o and g^u.

Key Predictions of the Model

Given that we have both randomly assigned teams and voluntarily formed teams, the model can already make some direct predictions on the choice of teammate characteristics.

Proposition 3.6. *(Match on Observables) If the quality channel of diversity is dominated by the direct cost of diversity, namely,* $\frac{\partial F}{\partial q}\frac{\mathrm{d}q}{\mathrm{d}g} - \frac{\partial C}{\partial d}\frac{\mathrm{d}d}{\mathrm{d}g} \leq 0$*, then,* $g^o \sim \mathcal{F}^N$ *will be first-order stochastically dominated by* $g^o \sim \mathcal{F}^X$

$$\forall \tilde{g}: \; \mathbb{P}_{\mathcal{F}^X}\{g^o < \tilde{g}\} \leq \mathbb{P}_{\mathcal{F}^N}\{g^o < \tilde{g}\}. \tag{3.12}$$

See appendix for proof. Compared to exogenously assigned teams, endogenously formed teams are more likely to match on observable characteristics, as long as the direct cost of diversity on utility is sufficiently large compared to any potential quality benefits.

Proposition 3.7. *(Match on Unobservables) If the quality channel of unobserved diversity is domi-*

[9] The notion of the positive correlation between observed and unobserved controls is motivated by the core assumptions in the selection correction literature of Altonji, Elder, and Taber (2005) and Oster (2019).

nated by the direct cost channel, namely, $\frac{\partial F}{\partial q}\frac{\partial q}{\partial g^u} - \frac{\partial C}{\partial d}\frac{\partial d}{\partial g^u} \leq 0$, then endogenously formed teams are more likely to match on unobservable characteristics than exogenously assigned teams, conditioning on the same level of observed diversity

$$\mathbb{E}_{\mathcal{F}^X}[g^u|g^o] \geq \mathbb{E}_{\mathcal{F}^N}[g^u|g^o]. \tag{3.13}$$

See appendix for proof. Conditional on g^o, endogenously formed teams are more likely to match on unobservable characteristics than exogenously assigned teams if the direct cost of diversity is high. Conversely, if the quality channel dominates, endogenously formed teams will have greater complementarity on unobservable skills than exogenously assigned teams instead. To clarify, in theory, one would not be able to formulate a test against true unobservables. However, given that the researchers do observe a wide array of characteristics, we can choose what to include in the conditioning set as "observable" and what to exclude from it as "unobservable".

Next, we derive the main prediction on how performance is related to observed diversity g^o.

Proposition 3.8. *(Performance Predictions on Observed Diversity) For any given distribution of team composition $\mathcal{F} \in \{\mathcal{F}^X, \mathcal{F}^N\}$, the sign on how performance F relates to observable diversity g^o is determined by the relative strength of the quality channel of observed diversity and the effort channel of observed diversity as follows*

$$\frac{\mathrm{d}}{\mathrm{d}g^o}\mathbb{E}\Big[F^*|g^o\Big]_{g\sim\mathcal{F}} = \frac{\mathrm{d}}{\mathrm{d}g^o}\mathbb{E}\Big[F(q(g^o,g^u),e^*(g^o,g^u))|g^o\Big]_{g\sim\mathcal{F}} \tag{3.14}$$

$$= \mathbb{E}\left[\underbrace{\frac{\partial F}{\partial q}\frac{\partial q}{\partial g^o} + \frac{\partial F}{\partial q}\frac{\partial q}{\partial g^u}\frac{\mathrm{d}g^u}{\mathrm{d}g^o}}_{\substack{\textit{The Quality Channel}\\ \textit{of Observed Diversity}}} + \underbrace{\frac{\partial F}{\partial e}\frac{\partial e^*}{\partial g^o} + \frac{\partial F}{\partial e}\frac{\partial e^*}{\partial g^u}\frac{\mathrm{d}g^u}{\mathrm{d}g^o}}_{\substack{\textit{The Effort Channel}\\ \textit{of Observed Diversity}}}\,\Bigg|\, g^o\right]_{g\sim\mathcal{F}}. \tag{3.15}$$

The result follows directly from Lemma 3.3. The proposition above describes the effort channel (or the quality channel) of observed diversity as the sum of the direct impact of observed diversity g^o on performance and the indirect impact on performance through its correlation with the unobserved component of diversity g^u.

Corollary 3.9. *(Weakened Performance Gradient under Endogenous Formation) The relationship between performance and observed diversity g^o is weaker under endogenous team formation than under* *exogenous team assignment*

$$\left|\frac{\mathrm{d}}{\mathrm{d}g^o}\mathbb{E}\Big[F^*|g^o\Big]_{g\sim\mathcal{F}^X}\right| > \left|\frac{\mathrm{d}}{\mathrm{d}g^o}\mathbb{E}\Big[F^*|g^o\Big]_{g\sim\mathcal{F}^N}\right|. \tag{3.16}$$

The result follows directly from Lemma 3.5, where the observed diversity g^o becomes a less informative measure of the true diversity g when students can select their teammates based on unobservable characteristics.

In summary, we have three key predictions based on the model. Proposition 3.6 and Proposition 3.7 provide predictions on the behavior of teammate choices, whereas Proposition 3.8 provides predictions on the relationship between performance and observed diversity, testing a model where the effort channel dominates from a model where the quality channel dominates.[10] Moreover, Corollary 3.9 is not a test against any model, but rather explains the differences in performance gradients regardless of the channel.

4 Data

We obtain anonymized administrative records on student characteristics and background of all matriculated MBA students from the class years 2013 to 2016. To complement official records, we also obtain self-reported student career and personal interests from an online student-service portal. We also obtain the team composition and performances in their FIELD 3 course.

Student Characteristics

We obtain anonymous data on student characteristics from the Office of MBA Student and Academic Services. We observe an extensive set of characteristics, including gender, race/ethnicity, home country,

[10]It is worthwhile to note that Proposition 3.6 and Proposition 3.7 provides the prediction to differentiate the quality channel of diversity $\frac{\partial F}{\partial q}\frac{\mathrm{d}q}{\mathrm{d}g}$ from the direct cost of diversity $\frac{\partial C}{\partial d}\frac{\mathrm{d}d}{\mathrm{d}g}$, whereas Proposition 3.8 tests the quality channel of diversity $\frac{\partial F}{\partial q}\frac{\mathrm{d}q}{\mathrm{d}g}$ against the *indirect* channel of diversity through reduced effort $\frac{\partial F}{\partial e}\frac{\mathrm{d}e^*}{\mathrm{d}g}$. While we could have written down a simpler model, such as $U(g) = F(q(g), d(g))$, where the optimal effort is abstracted away, we think it is useful to allow for a more explicit model of the optimal effort. There are several reasons: Anecdotally, we have learned that students teams of varying characteristics exert vastly different levels of efforts, which one could directly test if one had the requisite data. More importantly, to the extent that in practical settings such as delegated asset management, the agent typically does not fully internalize the performance benefit but does fully bear the cost, in which case, it is useful to allow for modeling the performance and the cost separately. Lastly, a more explicit model of effort is also useful in considering the impact of potential policy interventions that subsidize effort directly. That said, if we abstracted away the linkage through optimal effort, all three propositions would be testing the same simple model in terms of the comparison of the direct quality channel and the direct cost channel.

undergraduate institution, past employers, and past industry experience of each MBA student from the class years 2013 to 2016. We were not provided with students' actual names.

Table 1 reports summary statistics for the 3,684 MBA students in our sample. Women make up 41% of the total student population. Approximately 38% of the students are White, 12% are Asian American, 5% are Black, 4% are Hispanic, and 35% are international. India, Canada, and China represent the top three origin countries for international students, as shown in Appendix Table 1. In terms of past work experience, roughly half of the students worked in finance or consulting before business school. The big three consulting firms (McKinsey, Bain, and BCG) and bulge bracket investment banks (Goldman Sachs and Morgan Stanley) are the top five past employers for Harvard MBA students, as shown in Table 2. On average, 11% of students had experience in the technology industry. 25% of the MBA students graduated from Ivy League schools, and 38% of them graduated from a set of top 20 schools. Student characteristics are stable across class years.

To complement the administrative record, we also obtain complementary information about students' career and personal interests. Specifically, most MBA students maintain a "Class Card" as part of their online student profile that is accessible to other students. In the Class Card, they list biographic information as well as their career interests and club memberships. In terms of career interests, 80% of students report at least one career interest, and Appendix Table B.2 shows the most commonly reported ones include finance (48%), technology (35%), consulting (26%), and entrepreneurship (24%). In terms of club memberships, about 30% students list at least one membership affiliation, and there are 80 unique clubs reported in the data. Appendix Table B.3 shows that the most popular ones are the VCPE Club (26%), Entrepreneurship Club (19%), and Social Enterprise Club (16%). While many of the clubs are career-oriented, many of them could be considered social clubs (e.g., Basketball Club, Wine Club, and Christian Club).

Lastly, we collected the complete rosters of all MBA students who graduated with distinctions from graduation brochures. Students whose grades fall into the top 20% of their section in both their first and second years are given the award. Approximately 12% of MBA students in our sample graduate with distinctions.

Team Performance

Beyond student characteristics and team membership, we also collect information about team performance from the MBA program office, summarized in Table 3. Specifically, we code team outcomes into four binary indicators:

(i) *IPO Day*: We observe whether a team progresses to the IPO Day. Approximately 75% of the teams were determined to be sufficiently developed to present on IPO Day. Otherwise, it was placed in the "Failed Business Track."

(ii) *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".

(iii) *Section Top 3*: A project was ranked in the top 3 of their section by a panel of judges. Because a typical section has approximately 15 teams, about 20% of all projects were ranked as Section Top 3.

(iv) *Class Top 3*: A project was ranked as top 3 in the entire class. 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 did not progress, and the median quantile of this group is 0.125. All other categories are defined analogously: IPO Day Only (0.375), Viable (0.65), Section Top 3 (0.89), and Class Top 3 (0.99). Our performance measure is increasing in the project outcome.

While it may be a concern that a classroom setting may not provide sufficient incentives for the participating students to perform, we believe that the public nature of the final presentations, namely in front of their classmates and external VCs, provides additional incentives to do well beyond just grades. In our setting, unlike a typical classroom setting, the grades are de facto public, because the participants of the IPO days are known and the designation of being "viable", "section top 3" and "class top 3" are all announced as part of the IPO day outcomes. Thus, the combination of social pressure and the possibility of connecting to actual external investor beyond the classroom can provide additional incentives to perform.

5 Properties of Team Formation

In this section, we describe the properties of teams created under random assignment and voluntary formation. We first verify key properties of the conditional random assignment used for Class 2013 to ensure that randomization was performed properly. Next, we show that homophily plays a central role in how teams are formed when students choose their teammates voluntarily, where homophily along the lines of gender, race, ethnicity, education, and industry all play significant roles.

5.1 Conditional Random Assignment

For the Class of 2013, the assignment algorithm randomly created teams conditional on two specific student characteristics. Namely, the algorithm was developed to ensure that the composition of each team created by the computer approximately reflected the overall composition of the entire section in terms of gender and whether a student was from the US or international.

We first 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 1a shows the distribution of the number of female students on 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,[11] 62% had two women, and 38% had three women. There were zero teams with more than three women. There were also zero teams with no women. In sharp contrast, when students formed teams voluntarily in 2014 and 2015, out of 300 six-person teams, 12% of the teams had no women, 12% of the teams had one woman, 53% of the teams had 2 or 3 women, and another 19% of the teams had four or more women. Figure 1b shows 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 assignments, but 20% had over four international students with voluntary formation.

[11] 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.

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 endogenously in the next subsection. Specifically, we construct pseudo-student pairs by matching each student to every other student within the same section and year. This process creates 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/ethnicity, school, industry) tie equals one if two students belong to the same gender (race/ethnicity, school, industry) group.

Gender ties assume a value of one when two students are both coded as male or female by the MBA program office based on their application files and zero otherwise. During the sample period, there were no other gender markers provided. Race/ethnicity ties are defined as follows: For domestic students, the race and ethnicity groups are Whites, Asian Americans, Hispanics, and Blacks. For international students, we divide the globe into six 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 if two domestic students belong to the same racial or ethnic group, or if two international students hail from the same region.[12] School ties assume a value of one when two students graduated from the same undergraduate institution and zero otherwise. Industry ties are defined by the broad industry background based on a student's pre-MBA work experience, categorized into finance, consulting, technology, and others.

To illustrate, consider the following example: James Brown is a student in Section A, which has 90 students, and he needs to form a team of six. We create 89 student-student pairs by matching Mr. Brown to all his section mates, where each pair is a potential teammate with whom Mr. Brown could form a team. If the match happened randomly, Mr. Brown would pair with an arbitrary teammate with a probability of approximately 5.6% (5 out of 89). The outcome variable $real\_match$ equals 1 for the five pairs with whom Mr. Brown is matched in the same team. To measure factors that affect team formation, we compare the probability of being in the same team ($real\_match$) when a pair of

[12] Admittedly, there may be various ways to characterize race/ethnicity ties, and the strengths of the tie may well vary greatly by the race/ethnic group. As a result, for our main results, we provide additional results where we analyze domestic and foreign student ties separately and each race/ethnicity separately.

students share the same characteristics (e.g., gender, race/ethnicity, school, and industry) with the probability of matching when a student pair have different characteristics.

Indeed, Table 4 validates that, once controlling for 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 assignments in the subsamples of student pairs that share/do not share their international status. Hence, the patterns in the validation test are also consistent with our interviews of HBS program officers when they said team switching was prohibited for the Class of 2013 once teams were assigned by the compute algorithm.

5.2 Homophily in Team Formation

In this section, we show that homophily along the dimensions of gender, race/ethnicity, educational background, and past work experience all play a significant role in team formation when teams are formed voluntarily.

Although the homophily phenomenon of "birds of 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 test the presence of homophily as we can precisely observe one's *entire* choice set (i.e., all members of the section) and therefore are not confounded by unobserved networks. Another advantage of our setting is that we have a wide array of individual characteristics so that we can examine the impact of homophily in many dimensions. Finally, given the nature of the admissions and the sectioning algorithm, students almost universally come to HBS without prior personal relationships.

Using the pseudo pair specification described before, Table 5 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 percentage points. Given the base rate of matching is 5.6%, this represents a substantial increase of 25% from the baseline probability of randomly matching with a student from the same race/ethnicity. Similarly, we find that 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 5 column (2) reports the regression result using the 2013 subsample. Given that teams were randomly assigned, the coefficients on race/ethnicity, school, and industry ties are not statistically significantly different from zero. The negative matching coefficient in front of gender reflects HBS's gender-balanced assignment mechanism.[13]

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 Schultz, 2005). Alternatively, we find all the results remain statistically significant using a randomization inference procedure.[14]

5.3 Breakdown of Homophily by Subgroups

We further investigate the strength of homophily by subgroup. 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

[13]It is worthwhile to clarify that the null hypothesis here is that shared characteristics do not predict team formation, namely, a coefficient of zero. Thus, finding a positive coefficient should be interpreted as shared characteristics is predictive of team formation compared to random matching. Finding a negative coefficient is interpreted as shared characteristics *is* predictive of not being in the same team compared to random matching. The magnitude of the coefficients should be interpreted as the extent to which shared characteristics is predictive of team formation. It should not be directly interpreted as a measure of preferences, as we cannot separately identify preferences from constraints. For example, if we observed in the data that both students coming from Japan is very predictive of team formation, it could be because that Japanese students strongly prefer to work with other Japanese students, or there are very few Japanese. On the other hand, both students coming from the US may not be very predictive of team formation, but it could be because there are numerous possible US-US pairs to choose from, even if the preference to work with fellow Americans is just as strong for Americans. In other words, without further variations on choice sets, we do not readily identify underlying preferences.

[14]Specifically, we create 1,000 random permutations of alternative team formations and re-run the matching regression on the generated data. In Appendix Table B.6, 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.

to form teams within their group. Table 6, 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 an over 20% increase relative to the baseline of random matching. The coefficient for Black students is 1.3%, but we lack statistical power, likely because Black students only make up 5% of the student body. 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 B.7 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%, about twice as large as the coefficients for European and South Asian students. We interpret this result as suggestive evidence that language and cultural barriers may be important contributors to how teams are formed and, consequently, how performances are affected, which we turn to in the next section.

In terms of gender, Table 6 column (2) shows that both men and women exhibit homophilic tendencies in team formation. In Table 7, 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 find that homophily plays a central and significant role in how teams are formed in our setting. To the extent that such sociological forces are likely to be present outside of our specific setting, this suggests that homophily alone (beyond other existing frictions) may pose meaningful challenges for minorities to break into the current entrepreneurship ecosystem, which is predominantly White and male.

6 Performance Implications of Team Diversity

In this section, we analyze the impact of team diversity on performance.

6.1 Definition of Diversity Scores

Our unit of performance analysis is at the team level. We measure team diversity across various dimensions–race/ethnicity, gender, school, and industry–and construct the diversity measure for each dimension as the fraction of student-pairs with different characteristics:

$$\text{Diversity Score}_i = \frac{\text{Number of ties with different characteristics in team}_i}{\text{Number of total possible ties in team}_i}, \tag{6.1}$$

or equivalently,

$$\text{Diversity Score}_i = 1 - \frac{\text{Number of ties with shared characteristics in team}_i}{\text{Number of total possible ties in team}_i}, \tag{6.2}$$

where the second term can be thought of as a homogeneity score.

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 the Middle East. 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.

6.2 Model Predictions: Stochastic Dominance

With the observed diversity score g^o defined above, we can test the first model prediction as described in Proposition 3.6 regarding stochastic dominance.

Figure 2 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 homogeneous teams.

When plotted as a cumulative distribution function, Figure 2 Panel B shows a larger area under the curve for voluntarily formed teams than randomly assigned teams.

The average race/ethnicity diversity score decreased from 0.76 under random assignment in 2013 to 0.72 under voluntary formation in 2014-2016. The average gender diversity score 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 under voluntary team formation, as documented in the previous section.

The summary statistics on the diversity score show that the distribution of the observed diversity score under endogenous formation is stochastically dominated by the distribution under random assignment, or more simply, endogenously formed teams are more homogeneous. The result of this first test is that our data is consistent with a model where the direct cost of diversity dominates the quality channel.

6.3 Performance Gradient under Random Assignment

In this section, we examine the impact of diversity on team performance. As described in the data section, our main performance measure is the median quantile of the team's project, but our results are also robust to alternative measures such as the raw binary outcomes.

Graphically, Figure 3 shows the binscatter of team performance on diversity scores in race/ethnicity. Panel A on the top indicates that among randomly assigned teams (2013), higher diversity scores correspond to poorer team performance. Panel B on the bottom 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 8 summarizes the main performance results. Column (1) shows the coefficient on the race/ethnicity score is -0.48 for the 2013 cohort with randomly assigned teams. Since the standard deviation of the race/ethnicity score is 0.18,[15] it suggests that a one-standard-deviation increase in racial/ethnic diversity leads to approximately a 9 percentile decline in the performance rankings (e.g., a decline from being ranked at 80th percentile to 71st percentile), or, equivalently, an 18% decline in performance, as the median team is ranked at the 50th percentile. Thus, the magnitude of the negative impact

[15] 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.

of diversity on performance is both statistically and economically significant. Therefore, when teams were exogenously assigned, we find that racially/ethnically diverse teams performed worse than more homogeneous teams. Through the lens of our model, the negative coefficient suggests that the effort channel of observed diversity dominates any plausible quality channel, leading to a negative performance gradient, as described in Proposition 3.8.

Because the assignment of teams is randomized, these students have no ability to select their teammates based on either observable or unobservable student characteristics. Moreover, whether a given student will land in a diverse team or homogeneous team is completely exogenous to their own characteristics. 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 setting when teams are randomly assigned.

Our setting bypasses the usual empirical challenge associated with interpreting the correlation between diversity and performance. In a typical organizational environment, one may be concerned that more diverse teams are self-selected to be able to work well with each other along other dimensions. More diverse teams may also attract members who are unobservably better at collaborating with colleagues from different backgrounds.[16] Yet, our empirical estimate does not suffer from the usual upward bias associated with such selection on unobservables.

Diversity vs. Diversity Correlates

Even though the random assignment successfully removes the scope under which the team diversity may be subject to unobserved student selection, a potential issue in interpreting the regression result is whether the negative performance is caused by team diversity itself or characteristic correlates of it.

For instance, more diverse teams may have a higher fraction of Asian Americans, but we also know that in the data, Asian Americans are more likely to have experience in the technology sector (Appendix Table B.4), which may be predictive of their performance in the class. In other words, the presence of performance-relevant student characteristics that are also correlated with our diversity score can potentially affect how we interpret our result.

[16] Technically, selection on one's ability to collaborate is not explicitly captured by our theoretical model, but our model can readily incorporate it by adding another dimension of unobserved group characteristics.

Conceptually, the interpretation issue arises because the randomization takes place at the student level, as opposed to at the characteristics level. When the computer algorithm randomly assigns students into teams, it results in variations in the levels of racial/ethnic diversity. Other observed and unobserved characteristics (e.g., industry, education, language, academic quality, personality, risk preferences, etc.) of the team would all vary by team diversity because of the underlying, *unconditional* correlation between race/ethnicity and these characteristics.

By contrast, our research design is *not* randomly assigning student race/ethnicity while holding all other characteristics fixed. Although this would technically resolve the interpretation issue, we believe that this alternative research design is neither feasible nor relevant. It is simply not feasible to randomly assign race/ethnicity holding all other characteristics fixed, because one simply cannot alter their race/ethnicity without altering underlying characteristics that are associated with it (e.g., they would have to change their entire life experience growing up as a different race/ethnicity.) Moreover, it is also empirically irrelevant, for instance, when considering various diversity related policy in the form of hard or soft quotas.

As a result, given the randomization at the student level in our setting, we address the issue of interpretation by controlling for a variety of likely correlates and evaluate the empirical impact of them. In Table 8 column (2), we directly control for student characteristics that may be indicative of their quality. Specifically, we control for the percentage of students with prior startup experience, the percentage of students who came from top undergraduate institutions,[17] and the percentage of students graduating with MBA honor.[18] These variables are potential proxies for students' abilities. We observe that the results remain significant, and the magnitude of the coefficient stays very similar (-0.48 vs. -0.45).

In Table 9, we control for an even wide set of student characteristics that might have served as demographic correlates. It is remarkable that the main coefficient is barely changed by the inclusion

[17] We classify as top universities the Ivy League schools (Brown University, Columbia University, Cornell University, Dartmouth College, Harvard University, Princeton University, University of Pennsylvania, and Yale University) as well as other top schools (Amherst College, California Institute of Technology, Cambridge University, Duke University, MIT, Northwestern University, Stanford University, University of California, Berkeley, University of Chicago, University of Oxford, and Williams College).

[18] Ideally, we would like to measure a student's academic strength at the time when they took FIELD 3. However, in the absence of their college GPA or GMAT scores, we use honors status as a proxy because we believe academic achievements are likely serially correlated.

of such a wide set of controls, resulting in estimates in a narrow range from -0.48 to -0.50. Column (2) controls for the fraction of the team with prior finance, consulting, or technology experiences. Column (3) controls for additional performance correlates, such as the fraction of students who attended undergraduate institutions located in English speaking countries. Column (4) controls for key demographic characteristics including gender and race/ethnicity. In Appendix Table B.8, we also show a specification where all student characteristics controls are included. In such a specification, the statistical significance of the estimate is much reduced, yet the point estimate remains unaffected at -0.48.

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 had a substantial impact on the results gives us more confidence in interpreting the negative coefficient as driven by the level of team diversity as opposed to these observable demographic or quality correlates.

Drivers of Negative Performance

If not for these demographic or quality correlates, what might be the driver for the negative performance gradient?

In Table 10, we break down the diversity score into various sub-groups to investigate further the source of performance impact. In column (2), we break down our diversity score into a domestic component, namely, mismatched races/ethnicities among domestic students, and an international component, namely, mismatched regional origins among international students. We find that both components admit significant negative coefficients, whereby the diversity score among international students has an even larger negative coefficient. In column (3), we further break down the domestic component into different subgroups. We find negative coefficients for all subgroups (Whites, Asian Americans, Blacks, Hispanics), meaning that all their performances suffer when they are teamed up with students that are of a different race/ethnicity from their own. Similarly, in column (3), we further break down the international component into different regions. We find negative coefficients for most regions of the world (Europe, South Asia, East Asia, and Latin America), meaning that their performance suffers when working with someone from a different region of their own. For groups with small populations, we obtain weaker statistical significance, but almost all of them are negative.

Given the negative coefficient across almost all racial and ethnic subgroups, our interpretation is that simply having to work with someone of *different* race/ethnicity poses some challenges to the team production process, thus reducing overall performance.[19] Through the lens of our model, it is consistent with the notion that the effort channel of diversity dominates the quality channel of diversity. In other words, having to work with someone of a different race/ethnicity drives up the cost of effort, leading to less effective effort exerted and resulting in worse performances. Moreover, whatever potential skill complementarity that is brought about by diversity is not enough to overcome it.

6.4 Performance Gradient under Voluntary Formation

We estimate the correlation between performance and observed racial/ethnic diversity for the Classes of 2014 to 2016 when teams are formed voluntarily. Because team formation is entirely voluntary, students will be able to select on characteristics, some of them observable to the researcher and some unobservable. Naturally, the coefficients in front of the diversity score are mere correlations and do not assume any causal interpretation.

We find a much weaker negative correlation between team diversity and performance when students choose their own teammates. Table 8 columns (3) and (4) examine the correlation for the sub-sample of voluntarily formed teams. Column (3) shows that higher racial/ethnic diversity is still associated with poorer performances, but the coefficient (-0.18) is about only one-third of the magnitude compared to the randomly assigned teams (-0.48). Column (4) allows for more quality controls, as described before, illustrating that prior start-up experience and being an academically strong student lead to significantly better performance.

Crucially, even though the performance gradient estimated under voluntary formation is by no means causal, the *change* in the negative performance gradient due to voluntary formation is causal: namely, allowing students to choose their own teammates as opposed to exogenously assigning them leads to a reduction in the performance penalty of racial/ethnic diversity. Because it is unlikely any significant portion of students change their HBS matriculation decision based on the course design of

[19] As described in the literature (Hoogendoorn, Oosterbeek, and van Praag, 2013), a research design like ours do not and cannot pinpoint to the specific underlying mechanism of it (i.e., whether it is due to differences in language or differences in interests) because as all these characteristics are correlated with race/ethnicity and, thus, cannot be separately identified in a research design where the randomization takes place at the individual level rather than at the characteristics level.

one single course, we believe whether a student is subject to exogenous or endogenous assignment in their FIELD 3 course is likely orthogonal to their characteristics.

Table 8 columns (5) and (6) examine the full sample. The coefficient on the interaction term Voluntarily Formed × Race/Ethnicity Score is 0.30, suggesting that allowing students to choose their own teams significantly alleviates the performance penalty of diversity by approximately 60% (0.30/0.49). However, note that it is also at the cost of creating fewer diverse teams overall.

When controlling for more detailed demographic and quality correlates, we also find a weaker relationship between team diversity and performance among these endogenously-formed teams. In Table 9 columns (6) - (8), the coefficient on diversity remains negative but much smaller in magnitudes than the exogenously-assigned teams (columns (1)-(4)), and they sometimes lose statistical significance. When broken down further by race and ethnicity subgroups, Table 10 columns (7) - (10) produce much smaller and statistically insignificant coefficients, in contrast to the exogenously assigned teams (columns (1) - (5)).

Taken together, the much-weakened coefficients suggest that, even with the same level of *observed* diversity, there is likely a great deal of endogenous selection on unobservables that can overcome the cost associated with having to work across different races or ethnicities. Through the lens of the model, this suggests the correlation between the unobserved and observed characteristics $\frac{\mathrm{d}g^u}{\mathrm{d}g^o}$ must be sufficiently different between exogenous assignment $\mathcal{F}^X$ and endogenous formation $\mathcal{F}^N$. Therefore, the weakened relationship between team diversity and performance among endogenous teams is consistent with the prediction as described in Corollary 3.9, where the scope to select on characteristics likely played an important role.

The presence of both exogenously-assigned teams and endogenously-formed teams in an otherwise identical setting is a particularly novel and unique feature of our empirical setting. Typical field experiments would only analyze outcomes with randomized assignments, whereas typical observational studies would only include endogenously-formed teams. As such, we view our findings as unique in bridging the gap in interpreting the causally-estimated coefficients and observational correlations.

In terms of policy implication, while we are reluctant to draw overly broad conclusions, our findings does suggest the type of intervention used to achieve a certain level of diversity could have a significant impact on performance. For example, one takeaway could be that policy interventions aimed at improving "observed" diversity should allow for greater scopes of selection on unobservables to alleviate potential negative productivity impacts, for example, through a longer compliance horizon.

Matching on Unobservables

So, why does voluntarily formed diversity result in a much smaller performance penalty than randomly assigned diversity? Our hypothesis is that a voluntarily-formed team formation process allows individuals to select on many unobservable dimensions (e.g., career interests, personal interests, functional expertise, etc.) that can either reduce the cost of effort (e.g., through better communication and collaboration) or improve the quality of teams through complementary skills, conditioning on the same level of observable racial/ethnic diversity.

To test the hypothesis, in addition to the existing demographic data, we collected extensive additional data that may shed some light on such "unobservable" dimensions from their Class Card profiles, an internal student information portal. Indeed, Table 11 Panel A shows that, with voluntary team formation, a racially/ethnically mismatched student pair in a team is more likely to match on other dimensions such as their home state (column 1), the number of shared career interests (column 2), and the number of shared club memberships (column 3).[20] By contrast, Table 11 Panel B shows that a racially/ethnically mismatched student pair in a team is no more likely to select on complementary industry backgrounds, for example, measured in terms of the complementarity between finance and non-finance, the complementarity between tech and non-tech, or just having different industry backgrounds at all.[21]

Through the lens of the model, the findings in Table 11 maps to the second prediction of the model regarding matching on "unobservables" (Proposition 3.7), where we find suggestive evidence in

[20]Besides directly selecting on student characteristics, students could also be selecting on unobserved relationships, such as, whether they are already friends. Now, to the extent that friendship relationships might also be selecting on either shared or complementarity in characteristics in the first place, which we believe are highly likely, our model readily captures this channel, where we interpret the reduction in the cost of effort via friendship to ultimately still driven by selection on characteristics at an earlier stage of the game. Therefore, we view our model as general enough to accommodate selection on friendships during voluntary formation.

[21]We caution that we may be underpowered in detecting the presence of complementarity because we do not have detailed data on students' functional expertise. For example, both students may have previously worked in a tech firm, but one with sales and another one with engineering experience. Nonetheless, other parts of the model predictions still overwhelmingly suggest that the complementarity channel, even if present, will be dominated by the direct cost of diversity.

support of the greater alignment in unobservable characteristics among endogenously formed teams, conditioning on observed racial or ethnic differences, whereas we find little direct support for skill complementarity. Taken together with the predictions in the previous sections on stochastic dominance 3.6 and on the negative performance gradient 3.8, all three predictions point in the same direction in favor of a model where the quality channel is dominated by the direct cost of diversity or the indirectly cost of diversity through a reduced effort.

6.5 Robustness

In this subsection, we describe several robustness tests to our main results on the performance implications. We provide alternative specifications in terms of how team performance and team diversity are measured. We find all main results remain robust in these alternative specifications.

First, even though we use the median percentile as the performance measure, we could also perform our analysis using the underlying binary outcomes that indicate whether a team's performance is above a certain threshold, namely, IPO Day, Viable, and Section Top 3. Table 12 shows that diversity hampers performance at every stage of the project progression for the 2013 cohort: racially/ethnically more homogeneous teams are more likely to make it to the IPO Day presentation and to be considered viable. For Section Top 3, the results remain negative, but we lack the power to detect statistical significance. Moreover, consistent with our main results, when analyzed over the full sample, such negative performance implications among randomly assigned teams are also greatly alleviated when teams are formed voluntarily.

Next, given the underlying distribution of the diversity score is skewed to the left, as shown in Figure 2, 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 the skewness of the raw diversity score introduces additional non-linearity between performance and diversity, the variable transformation from diversity scores to diversity percentiles directly addresses this concern. We show in Table 13 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. In Appendix Figure A.1, we show that the relationship between performance and diversity percentiles remains robust, and they exhibit no noticeable non-linearity for both randomly assigned and voluntarily formed teams.

Lastly, we also explore other dimensions of diversity, such as diversity in gender, industry background, and prior education. In Table 14, we add measures of diversity in these additional dimensions. With these additional controls, our main result on race/ethnic diversity remains unchanged. In terms of gender diversity, because the team assignment software was designed by the course administrator to create gender-balanced teams for the Class of 2013, we, as researchers, regrettably are left with no variation in gender diversity to credibly estimate the impact of gender diversity on team performance. Consequently, we do not find any significant results on the gender score in any specifications.

In terms of diversity in prior education, Table 14 columns (5) and (6) both show a negative coefficient in the school score, which measures the fraction of team members who attended different undergraduate institutions, but with weaker statistical significance. Columns (5) and (6) also show that voluntary formation also seems to have alleviated the performance penalty in school diversity, although only significant at 10%. In terms of diversity in industry background, we do not find statically significant results, although all the signs are still consistent with our existing results, suggesting that diversity in observed industry background plays a relatively small role in our setting. Overall, despite much weaker statistical significance, it seems that mismatches in these other dimensions such as prior industry or school also degrade performance under exogenous assignment, and the relationship becomes weakened during endogenous formation, consistent with our main results in racial/ethnic diversity.

7 External Relevance

The highly structured nature of the FIELD 3 course at HBS allows us to exploit the exogenous as well as endogenous variation in the sample, allowing us to identify both the strong preferences to team up with those with shared characteristics as well as the strong performance implications of diversity in the 2013 cohort in which team membership is exogenous. However, it remains unclear whether the findings from our setting might be relevant more broadly, as there usually is a trade-off between broad empirical relevance and the ability to obtain well-identified results. In this section, we provide *suggestive* evidence of external relevance; i.e., can we find any evidence from real start-ups that provides insight as to whether our effects are present in a business setting?

First, we believe our empirical setting is relevant and realistic to the U.S. entrepreneurship ecosystem. Our sample population of MBA students at HBS is an integral part of the broader U.S. entrepreneurship network. The course set-up at Harvard Business School is intended for its MBA students; many of them have chosen to work in start-ups and the venture capital industry after graduation. Based on the exit surveys, over 20% of the graduating class entered the technology sector or venture capital in our sample. Many of them later progressed to leadership positions in their field. According to Gompers and Wang (2017), about 30% of all venture-backed founders in the U.S. have an MBA degree, among whom Harvard Business School ranks as *the* most representative institution, accounting for over 16% of all MBA founders, leading the next ones such as Stanford GSB (9%) and the Wharton School (6.7%) by a considerable margin. The performance measures of the proposed start-ups are evaluated realistically by relevant industry experts, comprised of a panel of experienced venture capital investors and entrepreneurs from the industry. Moreover, a handful of businesses that started during the course indeed continued to operate after the course, with some attracting significant outside funding.[22]

Next, we relate our findings to a similar set of matching and performance regressions using data from real founders. For this exercise, we examine all venture capital-backed US-domiciled start-ups from 1990 to 2016. Our data comes from VentureSource and Pitchbook. For each individual founder in the data, we hand-collect a broad range of biographic information, including gender, race, and ethnicity, through web searches, SEC filings, and news articles.

Founder genders are determined based on their first names. In the cases of unisex names, we determine gender by reading news articles and web pages mentioning or containing pictures of the individual founders. As for the ethnic background, we use the name-matching algorithm developed by Kerr and Lincoln (2010) to determine the most likely ethnicity of founders based on their last names. Individual founders are classified into five non-overlapping ethnic groups: East Asians, South Asians, Hispanics, and all others. To identify Black founders in our data, we manually examine pictures of all founders identified as White.

We determine the start-up outcome using VentureSource and Thomson Financial's SDC database,

[22]For example, Catalant, an online marketplace for consultants, was known to have started during FIELD 3 at HBS. https://en.wikipedia.org/wiki/Catalant.

supplemented by Thomson Financial's VentureXpert database. Although there are examples of successful investments that did not result in IPOs, the public floatation of a portfolio company is the cleanest signal of the start-up's success. We therefore consider a start-up to be successful if and only if it results in an IPO.

The majority of VC-backed start-ups in the U.S. are founded by a team as opposed to a single individual. Panel (a) tabulates the distribution of founding team sizes. 37.9% of start-ups have one founder, while 36.2% have two founders. 17.2% of the start-ups have founding teams of three people, while 6.0% have four founders. By comparison, our Field 3 setting has an average team size of six, perhaps resembling a combination of founders and the first few early employees.

We also find, consistent with the literature, that the demographics of VC-backed entrepreneurs in the U.S. are heavily skewed toward men, Whites, and Asians. Table 15 Panel (b) provides the summary statistics on the 26,401 founders during our sample period. Only 7.4% of the founders are women. 81% of founders are White, while 6.1% are East Asian and 9.1% are South Asian. Hispanic and Black founders are 3.3% and 0.3% of the sample.

Then, we examine the matching decision of founders based on race/ethnicity and gender. The unit of observation is a founder-founder pair. In addition to the real founder pairs, we create pseudo-pairs that match all founders in the same industry, same state, and same year. Real pairs are coded as 1 in the regression, and pseudo-pairs are coded as 0. Table 16 shows that having the same race/ethnicity increases the probability of matching by 0.86%. Compared to a baseline probability of 2.7%, this represents a 32% increase in the probability of matching relative to a random founder in the same year, industry, and state. All the individual races/ethnicities are positively and significantly related to forming a founding team together. Similarly, we find that if both founders are female, the probability of matching as founders is significantly higher by 2.3%, representing an over 100% increase in the probability of co-founding together compared to a random match.[23]

Finally, we look at the performance of the start-up based on the race/ethnicity score of the start-up team, calculated analogously as we did before in Eq (6.1). The dependent variable is whether the start-up has gone public. In Table 17, we find a negative, but statistically insignificant, relationship

[23] As before, the interpretation of the coefficients here is the extent to which shared gender/race/ethnicity can significantly predict the realized matches in the founder data.

between performance and observed racial/ethnic diversity. To the extent that the choice of whom to co-found a company is likely subject to a lot of deliberation by founders, the extent of selection on unobservable characteristics is likely quite large. The weak correlation between performance and racial/ethnic diversity in the start-up data appears consistent with our Field 3 results under voluntary team formation.

Overall, both the strength of homophily in founder selection and the performance slope of diversity using the actual founder data are consistent with our results from Field 3, which lend some support for the broader empirical relevance of our experimental setting. In terms of the relevance of the result under exogenous assignment, while it is hard to imagine a practical setting where members of a team have absolutely no control over who their teammates are, we argue that our exogenous assignment setting is especially useful in providing a "book-end", a kind of upper-bound on the extent of productivity loss due to "forced" diversity.

8 Conclusion

In this paper, we leverage various sources of randomization unique in our empirical setting to study the effects of diversity on team performance. We first document the ubiquitous presence of homophily in team formation, where shared ties in gender, ethnicity, education, and industry background all play an economically and statistically significant role. Then, when we examine the effect of team diversity on performance for two different types of team assignment mechanisms, we find that for teams in the 2013 cohort when team membership was exogenously assigned, greater racial and ethnic diversity leads to poorer team performance. However, when team membership was endogenously formed, such under-performance was much alleviated. Tying back to the stylized model, the empirical finding that diversity leads to worse team performance, together with auxiliary model predictions, is consistent with a model where the effort channel of diversity dominates the quality channel.

Moreover, beyond our specific setting, we show that our results are consistent with suggestive evidence from real startups. We find that in a sample of venture-backed startups from 1990 through 2016, founders are significantly more likely to team up with others who share gender and race/ethnicity. Also consistent with our Field 3 results, we find that the slope of the relationship between diversity and performance is negative, although, like the endogenous years of team selection in our setting, the coefficient is not significant.

Our results have important real-world implications. Our results on the performance effects of team 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 potential productivity loss 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 individuals could match on other characteristics that are not used by the computer algorithm, which in turn dampens the negative effect of diversity, such as shared career or personal interests, working styles, risk preferences, etc. Although we are not able to extract all of the information unobservable to the computer algorithm that has led to performance improvement under voluntary formation, we find some suggestive evidence that shared interests seem to matter. In this sense, we offer a lower bound and an upper bound on the performance gradient with respect to two extreme 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.

Overall, the results suggest that, to minimize potential adverse effects of diversity, policy interventions ought to take into account the process by which diverse teams are formed, especially in terms of selection on match-specific qualities. For example, it could take the form of allocating more resources to improve the matching process or by allowing for longer compliance timelines. In addition, to harness the full benefit of diversity, one could also target how to reduce the cost of effort required to work across different demographic lines. For example, subtle treatment effects may dislodge existing biases, where 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 performance improves after hiring. Overall, our findings pave a potential path to better achieve both diversity and performance goals.

References

Ahern, K. R., and A. K. Dittmar. 2012. The Changing of the Boards: The Impact on Firm Valuation of Mandated Female Board Representation. *The Quarterly Journal of Economics* 127:137–97.

Alesina, A., and E. L. Ferrara. 2005. Ethnic Diversity and Economic Performance. *Journal of Economic Literature* 43:762–800.

Altonji, J. G., T. E. Elder, and C. R. Taber. 2005. Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy* 151–84.

Aman-Rana, S., C. Minaudier, B. Alvarez Pereira, and S. Chaudry. 2021. Gender and Choice over Co-workers: Experimental Evidence. *SSRN Electronic Journal* .

Bandiera, O., I. Rasul, and I. Barankay. 2013. Team Incentives: Evidence from a Firm Level Experiment. *Journal of the European Economic Association* 11:1079–114.

Bernile, G., V. Bhagwat, and S. Yonker. 2018. Board diversity, firm risk, and corporate policies. *Journal of Financial Economics* 127:588–612.

Bertrand, M., S. E. Black, S. Jensen, and A. Lleras-Muney. 2019. Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway. *The Review of Economic Studies* 86:191–239.

Bertrand, M., and E. Duflo. 2017. Chapter 8 - Field Experiments on Discrimination. In A. V. Banerjee and E. Duflo, eds., *Handbook of Economic Field Experiments*, vol. 1 of *Handbook of Field Experiments*, 309–93. North-Holland.

Calder-Wang, S., and P. A. Gompers. 2021. And the children shall lead: Gender diversity and performance in venture capital. *Journal of Financial Economics* 142:1–22.

Calder-Wang, S., P. A. Gompers, K. Huang, and W. Levinson. 2023. Diversity in Venture Capital. In D. Cumming and B. Hammer, eds., *The Palgrave Encyclopedia of Private Equity*, 1–13. Cham: Springer International Publishing. ISBN 978-3-030-38738-9. doi:10.1007/978-3-030-38738-9_50-1.

Cohen, L., A. Frazzini, and C. Malloy. 2008. The Small World of Investing: Board Connections and Mutual Fund Returns. *Journal of Political Economy* 116:951–79.

Cook, L. D., M. Marx, and E. Yimfor. 2022. Funding Black High-Growth Startups. *NBER Working Paper* 30682.

Corwin, S. A., and P. Schultz. 2005. The Role of IPO Underwriting Syndicates: Pricing, Information Production, and Underwriter Competition. *The Journal of Finance* 60:443–86.

Currarini, S., M. O. Jackson, and P. Pin. 2009. An Economic Model of Friendship: Homophily, Minorities, and Segregation. *Econometrica* 77:1003–45.

Evans, R. B., M. P. Prado, A. E. Rizzo, and R. Zambrana. 2019. Identity, Diversity, and Team Performance: Evidence from U.S. Mutual Funds. *Working Paper* .

Ewens, M. 2022. Race and Gender in Entrepreneurial Finance. *NBER Working Paper* 30444.

Fairlie, R., A. Robb, and D. T. Robinson. 2022. Black and White: Access to Capital Among Minority-Owned Start-ups. *Management Science* 68:2377–400.

Gompers, P., J. Lerner, and D. Scharfstein. 2005. Entrepreneurial Spawning: Public Corporations and the Genesis of New Ventures, 1986 to 1999. *The Journal of Finance* 60:577–614.

Gompers, P. A., V. Mukharlyamov, and Y. Xuan. 2016. The cost of friendship. *Journal of Financial Economics* 119:626–44.

Gompers, P. A., and S. Q. Wang. 2017. Diversity in Innovation. *NBER Working Paper* 23082.

Gornall, W., and I. A. Strebulaev. 2021. The Economic Impact of Venture Capital: Evidence from Public Companies. *Working Paper* .

Hebert, C. 2020. Gender Stereotypes and Entrepreneur Financing. *Working Paper* .

Hjort, J. 2014. Ethnic Divisions and Production in Firms. *The Quarterly Journal of Economics* 129:1899–946.

Hochberg, Y. V., A. Ljungqvist, and Y. Lu. 2007. Whom You Know Matters: Venture Capital Networks and Investment Performance. *The Journal of Finance* 62:251–301.

———. 2010. Networking as a Barrier to Entry and the Competitive Supply of Venture Capital. *The Journal of Finance* 65:829–59.

Hoogendoorn, S., H. Oosterbeek, and M. van Praag. 2013. The Impact of Gender Diversity on the Performance of Business Teams: Evidence from a Field Experiment. *Management Science* 59:1514–28.

Hoogendoorn, S., and M. van Praag. 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.

Kerr, W., and W. Lincoln. 2010. The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention. *Journal of Labor Economics* 28:473–508.

Kleinbaum, A. M., T. E. Stuart, and M. L. Tushman. 2013. Discretion within constraint: homophily and structure in a formal organization. *Organization Science* 24:1316+–.

Lazear, E. P. 1999. Globalisation and the Market for Team-Mates. *The Economic Journal* 109:C15–40.

Lerner, J., and U. Malmendier. 2013. With a Little Help from My (Random) Friends: Success and Failure in Post-Business School Entrepreneurship. *The Review of Financial Studies* 26:2411–52.

Lu, Y., N. Y. Naik, and M. Teo. 2024. Diverse Hedge Funds. *The Review of Financial Studies* 37:639–83. Publisher: Oxford Academic.

Marsden, P. V. 1987. Core Discussion Networks of Americans. *American Sociological Review* 52:122–31.

———. 1988. Homogeneity in confiding relations. *Social Networks* 10:57–76.

Marx, B., V. Pons, and T. Suri. 2021. Diversity and team performance in a Kenyan organization. *Journal of Public Economics* 197:104332–.

Matsa, D. A., and A. R. Miller. 2013. A Female Style in Corporate Leadership? Evidence from Quotas. *American Economic Journal: Applied Economics* 5:136–69.

McPherson, M., L. Smith-Lovin, and J. M. Cook. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* 27:415–44.

Nanda, R., and J. B. Sørensen. 2010. Workplace Peers and Entrepreneurship. *Management Science* 56:1116–26.

Oster, E. 2019. Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics* 37:187–204. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/07350015.2016.1227711.

Prat, A. 2002. Should a team be homogeneous? *European Economic Review* 46:1187–207.

Reagans, R. 2011. Close encounters: analyzing how social similarity and propinquity contribute to strong network connections. *Organization Science* 22:835+–.

Ruef, M., H. E. Aldrich, and N. M. Carter. 2003. The Structure of Founding Teams: Homophily, Strong Ties, and Isolation among U.S. Entrepreneurs. *American Sociological Review* 68:195–222.

Shue, K. 2013. Executive Networks and Firm Policies: Evidence from the Random Assignment of MBA Peers. *The Review of Financial Studies* 26:1401–42.

Stolper, O., and A. Walter. 2018. Birds of a Feather: The Impact of Homophily on the Propensity to Follow Financial Advice. *The Review of Financial Studies* .

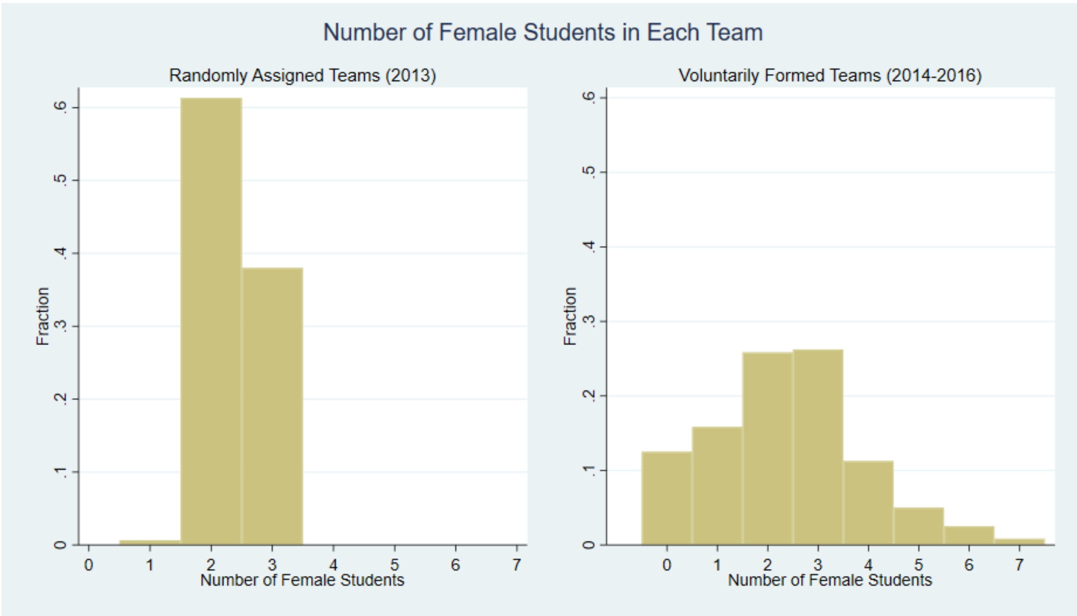
Sufi, A. 2007. Information Asymmetry and Financing Arrangements: Evidence from Syndicated Loans. *The Journal of Finance* 62:629–68.

van Knippenberg, D., and M. C. Schippers. 2007. Work Group Diversity. *Annual Review of Psychology* 58:515–41.

Figures

Figure 1: Distribution of Student Characteristics across Team Assignment Mechanisms

(a) The Number of Female Students within a Team



(b) The Number of International Students within a Team

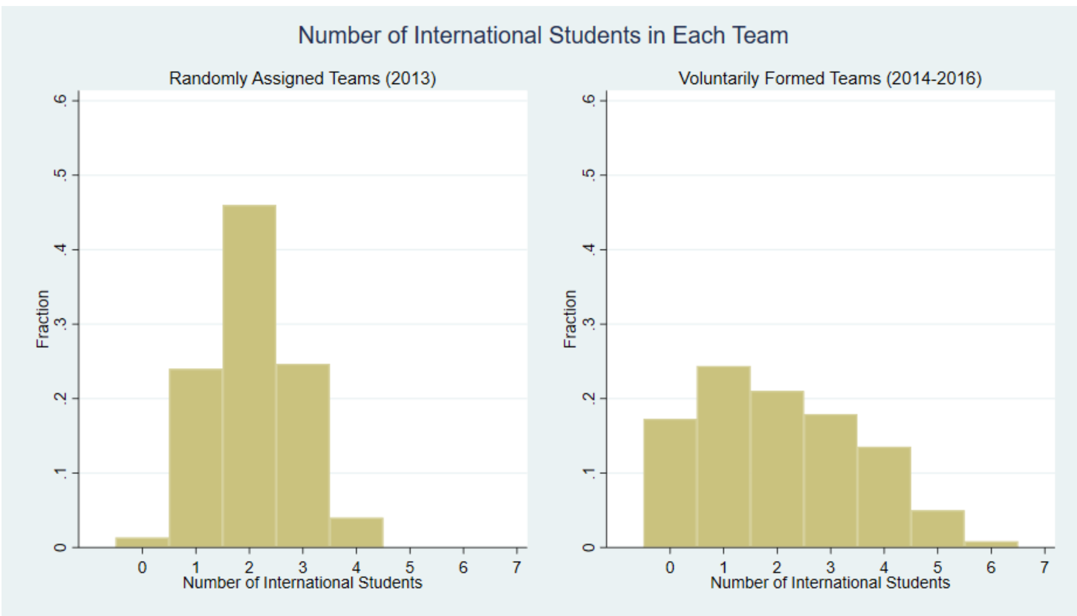
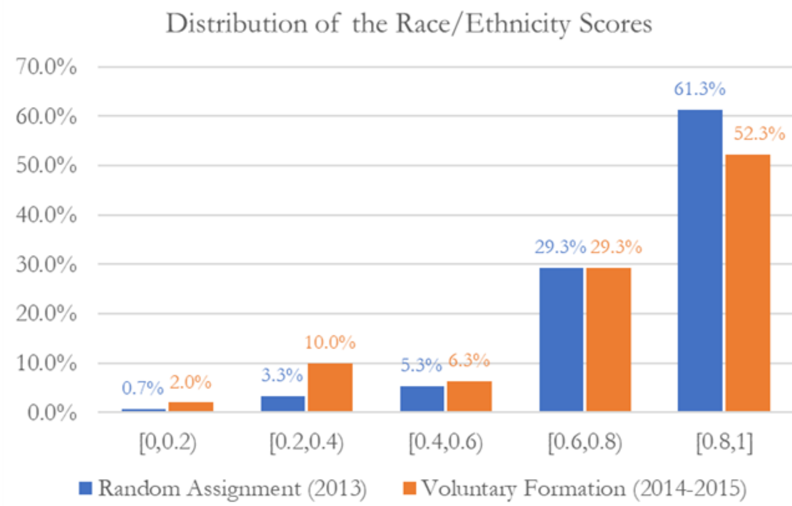


Figure 2: Distribution of Diversity Scores across Team Assignment Mechanisms

The figures below plot the probability distribution and the cumulative distribution of race/ethnicity scores under different team assignment mechanisms. The race/ethnicity diversity score is defined as the fraction of mismatched race or ethnicity ties within a team. Lower scores represent more homogeneous teams, whereas higher scores represent more diverse teams.

(a) The Distribution of Race/Ethnicity Scores



(b) The Cumulative Distribution of Race/Ethnicity Scores

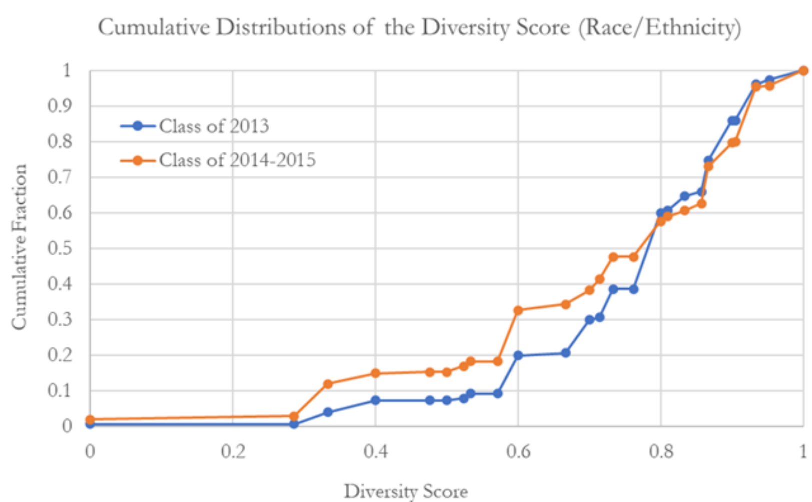
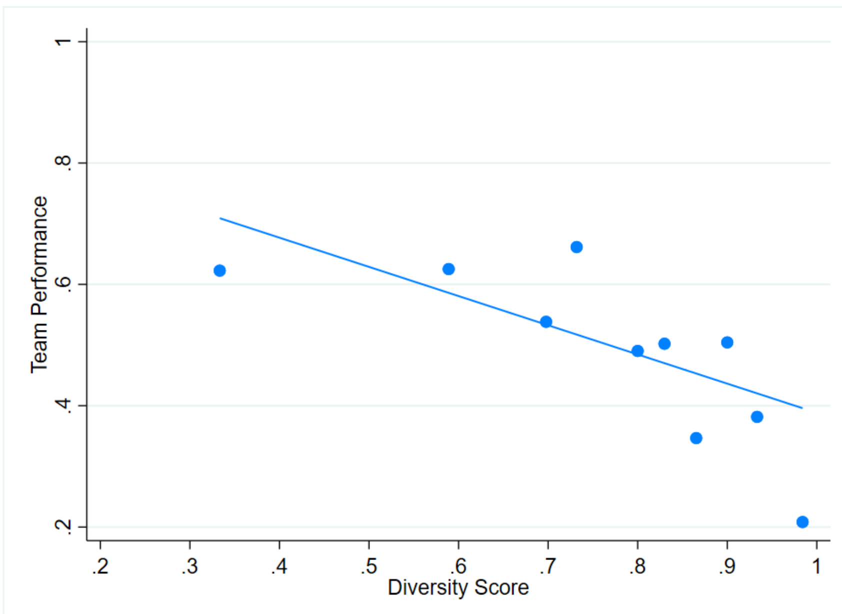


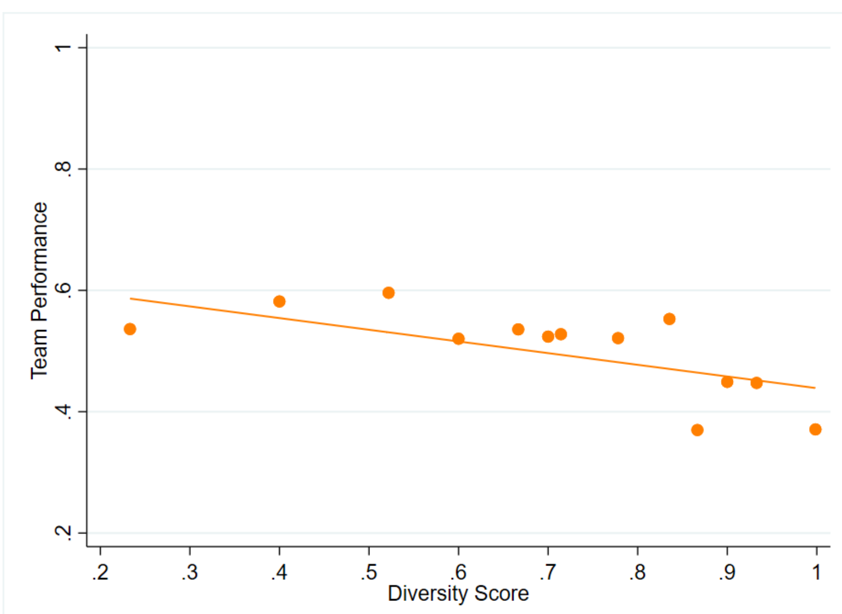
Figure 3: Racial/Ethnic Diversity and Team Performance

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 diversity score of the team, measured as the fraction of mismatched race or ethnicity ties within a team. Lower scores represent more homogeneous teams, whereas higher scores represent more diverse teams. The top panel plots the relationship between diversity and performance when teams are assigned randomly in 2013; the bottom panel plots the relationship when teams are formed voluntarily in 2014-2016.

(a) Diversity vs. Performance with Random Team Assignment (2013)



(b) Diversity vs. Performance with Voluntary Team Formation (2014-2016)



Tables

Table 1: Summary Statistics of MBA Backgrounds

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

	2013	2014	2015	2016	Total
# of Students	907	915	931	931	3684
Team Size	6.1	6.1	6.2	5.2	5.9
Age	28.9	29.1	29.1	29.2	29.1
% Female	39.3%	40.4%	41.1%	41.4%	40.6%
% White	37.2%	39.5%	37.7%	39.5%	38.5%
% Asian American	14.3%	11.8%	11.9%	11.8%	12.5%
% Black	4.5%	5.7%	5.6%	5.8%	5.4%
% Hispanic	3.7%	4.3%	4.8%	3.7%	4.1%
% International	34.1%	34.3%	34.6%	37.1%	35.0%
% Finance	29.7%	29.3%	33.8%	36.8%	32.4%
% Consulting	21.9%	20.5%	20.6%	25.1%	22.1%
% Technology	9.0%	9.8%	10.8%	14.0%	10.9%
% Healthcare	8.2%	7.9%	6.3%	8.9%	7.8%
% Undergraduate - Ivy League	26.9%	25.0%	23.6%	23.0%	24.6%
% Undergraduate - Top School	41.2%	37.9%	38.3%	34.3%	37.9%

Notes: Top schools are defined as all Ivy League schools plus MIT, Stanford, UChicago, Northwestern, UC Berkeley, CalTech, Duke, Williams College, Amherst College, Camdbridge University and Oxford University.

Table 2: Past Employment and Education Background

This table summarizes the employment and education background of MBA students prior to business school.

(a) Employment Background

Rank	Employer	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	3665	

(b) Education Background

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	Massachusetts Institute of Technology	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%
14	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	3500	

Table 3: 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 Mean	Performance StdEv
2013	150	78.7%	46.7%	20.0%	2.7%	0.50	0.28
2014	150	70.0%	39.3%	20.0%	2.0%	0.46	0.29
2015	150	73.3%	55.3%	20.0%	2.0%	0.51	0.29
2016	180	76.1%	52.8%	16.7%	2.2%	0.50	0.27
Total	630	74.6%	48.7%	19.0%	2.2%	0.50	0.28

Table 4: Properties of Conditional Random Assignment (Class of 2013)

This table summarizes the properties of the conditional random assignment algorithm used for team assignment for the Class of 2013. It reports the regression results of matching on various ties (race/ethnicity, gender, school, and industry ties) across all potential student pairs, conditioning on gender and international-status. 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. Statistical significance at the 1%, 5%, and 10% levels are denoted by ***, **, and *, respectively.

	Dependent Variable: Real Match					
	(1) Full Sample	(2) Same Gender	(3) Different Gender	(4) Both US	(5) Both Non-US	(6) US, Non-US Pairs
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)
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)			
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	81368	42140	39228	35228	6764	39376
R-Squared	0.002	0.000	0.000	0.002	0.003	0.001
Year FE	Y	Y	Y	Y	Y	Y

Table 5: Matching Regression

This table reports the regression results of matching on race/ethnicity, gender, education, and 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, education, industry) tie equals one if the pair has the same race/ethnicity (gender, education, industry). 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). Statistical significance at the 1%, 5%, and 10% levels are denoted by ***, **, and *, respectively.

	Dependent Variable: Real Match	
	Voluntarily Formed (2014-2016)	Randomly Assigned (2013)
	(1)	(2)
Race/Ethnicity Tie	0.014***	-0.00084
	(0.001)	(0.002)
Gender Tie	0.013***	-0.017***
	(0.001)	(0.001)
Both Same School	0.0085**	-0.0028
	(0.004)	(0.006)
Industry Tie	0.0062***	-0.00027
	(0.001)	(0.002)
Team Member Count	0.011***	0.011***
	(0.000)	(0.000)
Constant	-0.019***	0.00088
	(0.001)	(0.001)
Observations	254318	81368
R-Squared	0.003	0.001
Year FE	Y	Y

Table 6: Matching Regression: By Demographic 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, which equal one if both students share the same race/ethnicity or gender. For international students, we divide their home counties into six regions: Europe, East Asian, South Asian, Middle East, Africa, and Latin America. Two students form a tie if they hail from the same region. Robust standard errors are clustered at the student level. Statistical significance at the 1%, 5%, and 10% levels are denoted by ***, **, and *, respectively.

	Dependent Variable: Real Match			
	Voluntarily Formed (2014-2016)		Randomly Assigned (2013)	
	(1)	(2)	(3)	(4)
Both White	0.012***		-0.00024	
	(0.001)		(0.002)	
Both Asian American	0.014***		0.0019	
	(0.004)		(0.005)	
Both Hispanic	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	254318	254318	81368	81368
R-Squared	0.002	0.002	0.000	0.002
Year FE	Y	Y	Y	Y

Table 7: Matching Regression: By 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, which equal one if both students share the same education or industry background. Robust standard errors are clustered at the student level. Statistical significance at the 1%, 5%, and 10% levels are denoted by ***, **, and *, respectively.

	Dependent Variable: Real Match			
	Voluntarily Formed (2014-2016)		Randomly Assigned (2013)	
	(1)	(2)	(3)	(4)
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	254318	254318	81368	81368
R-Squared	0.001	0.001	0.000	0.000
Year FE	Y	Y	Y	Y

Table 8: 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 errors are clustered at year-section level. Statistical significance at the 1%, 5%, and 10% levels are denoted by ***, **, and *, respectively.

	Dependent Variable: Performance					
	Randomly Assigned (2013)		Voluntarily Formed (2014-2016)		Full Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
Race/Ethnicity Score	-0.48***	-0.45***	-0.18***	-0.13**	-0.48***	-0.46***
	(0.110)	(0.087)	(0.054)	(0.058)	(0.108)	(0.094)
Race/Ethnicity Score × Voluntary Formation					0.30**	0.32***
					(0.122)	(0.111)
Start-up Ratio		0.46		0.36**		0.35***
		(0.370)		(0.136)		(0.125)
Honor Student Ratio		0.32		0.25***		0.27***
		(0.175)		(0.089)		(0.077)
Top School Ratio		0.12		0.041		0.076
		(0.069)		(0.059)		(0.048)
Team Member Count	0.047	0.036	0.088***	0.073***	0.089***	0.073***
	(0.059)	(0.063)	(0.026)	(0.027)	(0.025)	(0.025)
Constant	0.59	0.52	0.045	0.043	0.33*	0.33**
	(0.358)	(0.365)	(0.165)	(0.157)	(0.169)	(0.155)
Observations	150	150	480	480	630	630
R-Squared	0.087	0.135	0.049	0.083	0.063	0.100
Year FE	N/A	N/A	Y	Y	Y	Y

Table 9: Impact of Team Diversity on Performance - Additional Controls

This table regresses team performance on team diversity scores controlling for a variety of student characteristics. 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 error is clustered at year-section level. Statistical significance at the 1%, 5%, and 10% levels are denoted by ***, **, and *, respectively.

	Dependent Variable: Performance							
	Randomly Assigned (2013)				Voluntarily Formed (2014-2016)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Race/ethnicity Score	-0.48***	-0.49***	-0.50***	-0.49**	-0.18***	-0.19***	-0.041	-0.078
	(0.110)	(0.128)	(0.145)	(0.215)	(0.054)	(0.058)	(0.071)	(0.091)
Finance %		0.030				-0.053		
		(0.164)				(0.069)		
Consulting %		0.083				-0.054		
		(0.168)				(0.082)		
Technology %		0.14				0.23**		
		(0.112)				(0.104)		
Start-up Ratio			0.49				0.34**	
			(0.391)				(0.134)	
Top School Ratio			0.12				-0.0088	
			(0.071)				(0.062)	
Honor Student Ratio			0.32*				0.26***	
			(0.173)				(0.092)	
English Speaker %			-0.12				0.14*	
			(0.164)				(0.071)	
% Female				-0.0012				0.13**
				(0.362)				(0.050)
% Black				-0.49*				0.00012
				(0.246)				(0.129)
% Asian				0.058				-0.012
				(0.111)				(0.079)
% Latino				-0.11				0.31**
				(0.169)				(0.135)
% White				-0.072				0.14
				(0.180)				(0.088)
Team Member Count	0.047	0.048	0.037	0.091	0.088***	0.089***	0.076***	0.089***
	(0.059)	(0.056)	(0.063)	(0.071)	(0.026)	(0.029)	(0.026)	(0.025)
Constant	0.59	0.55	0.64	0.37	0.045	0.057	-0.12	-0.15
	(0.358)	(0.352)	(0.407)	(0.470)	(0.165)	(0.174)	(0.181)	(0.189)
Observations	150	150	150	150	480	480	480	480
R-Squared	0.087	0.092	0.138	0.112	0.049	0.065	0.094	0.079
Year FE	N/A	N/A	N/A	N/A	Y	Y	Y	Y

Table 10: Impact of Team Diversity on Performance - Detailed Race/Ethnicity Groups

	Dependent Variable: Performance									
	Randomly Assigned (2013)					Voluntarily Formed (2014-2016)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Race/Ethnicity Score	-0.45***					-0.13**				
	(0.087)					(0.058)				
Race/Ethnicity Score (US Only)		-0.48***		-0.49***			-0.13**		-0.11*	
		(0.094)		(0.097)			(0.057)		(0.058)	
White American Score			-0.50***		-0.51***			-0.12**		-0.11*
			(0.097)		(0.102)			(0.057)		(0.058)
Asian American Score			-1.03***		-1.09**			0.24		0.27
			(0.302)		(0.338)			(0.246)		(0.245)
African American Score			-0.36		-0.12			-0.043		-0.026
			(1.687)		(1.599)			(0.760)		(0.765)
Hispanic American Score			-2.59		-3.66***			-0.95		-0.88
			(1.913)		(0.702)			(1.031)		(1.014)
Race/Ethnicity Score (Non-US)		-1.66***	-1.69***				0.26	0.28		
		(0.421)	(0.440)				(0.179)	(0.177)		
European Score				-1.86**	-1.89**				0.20	0.22
				(0.695)	(0.725)				(0.164)	(0.168)
South Asian Score				-2.66***	-2.99***				1.19***	1.22***
				(0.801)	(0.711)				(0.422)	(0.423)
East Asian Score				-0.42	-0.63				0.31	0.35
				(0.380)	(0.388)				(0.373)	(0.368)
Middle Eastern Score				0.24	1.59				0.73	0.76*
				(0.653)	(1.356)				(0.457)	(0.446)
Latin American Score				-0.84	-0.77				-0.16	-0.15
				(1.050)	(0.786)				(0.292)	(0.286)
Start-up Ratio	0.46	0.46	0.42	0.37	0.30	0.36**	0.35**	0.36**	0.34**	0.34**
	(0.370)	(0.370)	(0.375)	(0.409)	(0.425)	(0.136)	(0.139)	(0.138)	(0.143)	(0.141)
Top School Ratio	0.12	0.12	0.090	0.12	0.077	0.041	0.015	0.025	0.013	0.025
	(0.069)	(0.076)	(0.082)	(0.083)	(0.090)	(0.059)	(0.060)	(0.061)	(0.060)	(0.061)
Honor Student Ratio	0.32	0.27	0.26	0.27	0.26	0.25***	0.25**	0.24**	0.25**	0.24**
	(0.175)	(0.167)	(0.173)	(0.174)	(0.179)	(0.089)	(0.090)	(0.091)	(0.092)	(0.092)
Team Member Count	0.036	0.039	0.057	0.035	0.055	0.073***	0.076***	0.078***	0.076***	0.078***
	(0.063)	(0.062)	(0.070)	(0.062)	(0.071)	(0.027)	(0.026)	(0.026)	(0.024)	(0.025)
Constant	0.52	2.19***	6.11	6.10**	10.0*	0.043	-0.21	0.51	-2.22**	-1.74
	(0.365)	(0.651)	(4.091)	(2.123)	(5.108)	(0.157)	(0.246)	(1.350)	(1.083)	(1.674)
Observations	150	150	150	150	150	480	480	480	480	480
R-Squared	0.135	0.155	0.168	0.164	0.184	0.083	0.092	0.097	0.103	0.109
Year FE	N/A	N/A	N/A	N/A	N/A	Y	Y	Y	Y	Y

Table 11: Endogenous Selection of Teammate Characteristics

These two tables report the results of regressing the probability of match or mismatch with respect to certain student characteristics on the team formation mechanism, conditional on being mis-matched on race/ethnicity. The top panel (a) reports whether a student pair in a team with different race/ethnicity are more likely to match on certain other characteristics (e.g., home state, career interests, extracurricular activities) when they choose each other voluntarily. Home states are restricted to U.S. students only. There are 2,980 students (83%) who reported career interests in their class profile, and 1,102 students (31%) who report their extracurricular activities (e.g., sports, professional, and social clubs). The matching regressions are based on the sub-sample in which such additional information is available to both students in a pair. The bottom panel (b) reports whether a student pair in a team with different race/ethnicity are more likely to have complementary industry backgrounds (e.g., different backgrounds, finance and non-finance, tech and non-tech) when they choose each other voluntarily. Robust standard error is clustered at the student level. Statistical significance at the 1%, 5%, and 10% levels are denoted by ***, **, and *, respectively.

(a) Selection on Student Characteristics

	(1) Same Home State (US Only)	(2) Number of Shared Career Interests	(3) Number of Shared Extracurricular Clubs
Different Race/Ethnicity × Voluntary Formation	0.023*	0.17***	0.16**
	(0.013)	(0.063)	(0.082)
Different Race/Ethnicity	-0.0074	-0.12**	-0.12*
	(0.011)	(0.057)	(0.069)
Team Member Count	-0.011	0.038	0.073*
	(0.007)	(0.032)	(0.044)
Constant	0.14***	0.90***	-0.020
	(0.042)	(0.191)	(0.269)
Observations	11794	11896	1774
R-Squared	0.001	0.013	0.015
Year FE	Y	Y	Y

(b) (Lack of) Selection on Skill Complementarity

	(1) Different Industry Backgrounds	(2) Complementarity Finance and Non-Finance	(3) Complementarity Tech and Non-Tech
Different Race/Ethnicity × Voluntary Formation	-0.011	-0.011	-0.011
	(0.017)	(0.021)	(0.016)
Different Race/Ethnicity	-0.0019	-0.017	0.054***
	(0.015)	(0.018)	(0.014)
Team Member Count	-0.045***	0.027***	-0.0068
	(0.010)	(0.010)	(0.011)
Constant	1.11***	0.27***	0.16**
	(0.060)	(0.063)	(0.069)
Observations	18076	18076	18076
R-Squared	0.010	0.004	0.007
Year FE	Y	Y	Y

Table 12: Robustness: Impact of Team Diversity on Performance - Binary Outcomes

This table reports logistic regression results on the effect of race/ethnicity-Gender score. The dependent variables IPO day/Viable/Section 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 independent variables are diversity scores described in the paper. Robust standard errors are clustered at year-section level. Statistical significance at the 1%, 5%, and 10% levels are denoted by ***, **, and *, respectively.

	Randomly Assigned (2013)			Voluntarily Formed (2014-2016)			Full Sample		
	(1) IPO Day	(2) Viable	(3) Section Top 3	(4) IPO Day	(5) Viable	(6) Section Top 3	(7) IPO Day	(8) Viable	(9) Section Top 3
Race/Ethnicity Score	-4.71***	-4.44***	-1.27	-1.29**	-0.83*	-0.68	-4.83***	-4.45***	-1.34
	(1.407)	(1.259)	(1.044)	(0.522)	(0.466)	(0.521)	(1.445)	(1.221)	(1.024)
Race/Ethnicity Score × Voluntary Formation							3.59**	3.64***	0.67
							(1.507)	(1.298)	(1.170)
Start-up Ratio	3.03	3.01	3.54	2.14*	2.73***	2.93***	2.20*	2.78***	2.88***
	(3.279)	(2.126)	(3.524)	(1.273)	(1.031)	(1.124)	(1.177)	(0.930)	(1.084)
Honor Student Ratio	1.93	2.82*	1.78	1.28	1.63**	2.37***	1.42*	1.88***	2.22***
	(2.569)	(1.581)	(1.325)	(0.856)	(0.645)	(0.844)	(0.820)	(0.611)	(0.718)
Top School Ratio	1.31	0.58	0.77	0.57	0.42	-0.30	0.68	0.46	-0.044
	(0.867)	(0.417)	(0.987)	(0.545)	(0.363)	(0.589)	(0.466)	(0.298)	(0.500)
Team Member Count	0.97	0.25	-0.32	0.43*	0.69***	0.26	0.47**	0.64***	0.18
	(0.695)	(0.582)	(0.645)	(0.254)	(0.249)	(0.231)	(0.236)	(0.228)	(0.216)
Constant	-1.64	1.04	0.78	-1.30	-4.57***	-2.81*	1.81	-1.12	-1.86
	(3.911)	(3.511)	(3.256)	(1.472)	(1.483)	(1.575)	(1.734)	(1.654)	(1.359)
Observations	150	150	150	480	480	480	630	630	630
Year FE	N/A	N/A	N/A	Y	Y	Y	Y	Y	Y

Table 13: Robustness: Impact of Team Diversity on Performance - Normalized Scores

This table regresses team performance on normalized diversity scores measured by percentiles. The dependent variable Performance is the median of the quantile of the team's outcome. The independent variable is normalized diversity score, which is measured as the percentile of the underlying score in the simulated distribution of scores under total random assignment of teams. 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 1000 times. Robust standard error is clustered at year-section level. The statistical significance at the 1%, 5%, and 10% levels are denoted by ***, **, and *, respectively.

	Dependent Variable: Performance					
	Randomly Assigned (2013)		Voluntarily Formed (2014-2016)		Full Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
Race/Ethnicity Score Percentile	-0.34***	-0.31***	-0.15***	-0.13***	-0.34***	-0.32***
	(0.060)	(0.048)	(0.043)	(0.045)	(0.059)	(0.052)
Race/Ethnicity Score Percentile × Voluntary Formation					0.19**	0.19***
					(0.073)	(0.069)
Start-up Ratio		0.44		0.37***		0.37***
		(0.354)		(0.132)		(0.122)
Honor Student Ratio		0.27		0.25***		0.26***
		(0.173)		(0.089)		(0.078)
Top School Ratio		0.11		0.032		0.050
		(0.065)		(0.060)		(0.048)
Team Member Count	0.046	0.037	0.083***	0.069**	0.079***	0.064**
	(0.055)	(0.059)	(0.026)	(0.027)	(0.024)	(0.024)
Constant	0.39	0.34	0.020	0.041	0.19	0.21
	(0.322)	(0.344)	(0.163)	(0.156)	(0.146)	(0.138)
Observations	150	150	480	480	630	630
R-Squared	0.117	0.155	0.057	0.091	0.070	0.105
Year FE	N/A	N/A	Y	Y	Y	Y

Table 14: Robustness: Impact of Team Diversity on Performance - Multiple Scores

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					
	Randomly Assigned (2013)		Voluntarily Formed (2014-2016)		Full Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
Race/Ethnicity Score	-0.49***	-0.45***	-0.18***	-0.14**	-0.49***	-0.46***
	(0.105)	(0.088)	(0.056)	(0.059)	(0.103)	(0.091)
Gender Score	-0.025	-0.18	-0.043	-0.054	-0.00037	-0.14
	(0.658)	(0.598)	(0.059)	(0.059)	(0.630)	(0.582)
School Score	-0.96*	-0.75	0.20	0.33	-0.97**	-0.77
	(0.476)	(0.528)	(0.295)	(0.310)	(0.455)	(0.485)
Industry Score	-0.084	-0.048	0.12	0.12	-0.085	-0.048
	(0.165)	(0.165)	(0.082)	(0.084)	(0.156)	(0.153)
Race/Ethnicity Score × Voluntary Formation					0.30**	0.32***
					(0.117)	(0.107)
Gender Score × Voluntary Formation					-0.043	0.083
					(0.633)	(0.584)
School Score × Voluntary Formation					1.16**	1.10*
					(0.541)	(0.562)
Industry Score × Voluntary Formation					0.20	0.16
					(0.177)	(0.176)
Top School Ratio		0.060		0.064		0.063
		(0.067)		(0.061)		(0.050)
Start-up Ratio		0.46		0.34**		0.36***
		(0.365)		(0.136)		(0.125)
Honor Student Ratio		0.31		0.26***		0.27***
		(0.182)		(0.086)		(0.077)
Team Member Count	0.054	0.036	0.094***	0.078***	0.090***	0.074***
	(0.057)	(0.061)	(0.027)	(0.028)	(0.025)	(0.025)
Constant	1.58*	1.43	-0.26	-0.39	1.36*	1.21*
	(0.844)	(0.891)	(0.313)	(0.301)	(0.682)	(0.638)
Observations	150	150	480	480	630	630
R-Squared	0.107	0.145	0.056	0.091	0.067	0.103
Year FE	N/A	N/A	Y	Y	Y	Y

Table 15: Venture-Backed Founder Characteristics (Venture Source)

In this table, we analyze all venture-backed start-up teams domiciled the US from 1990 to 2016. In panel (a), we report founder characteristics. Founder ethnicity is identified using their last names augmented by extensive manual check of their profile pictures. Panel (b) reports the distribution of start-up team sizes.

(a) Summary Statistics on the Size of Founding Teams

Founding Team Size	Count	Percent (%)
1	4990	37.9
2	4770	36.2
3	2265	17.2
4	784	6.0
5 or More	356	2.7

(b) Summary Statistics on Venture-Backed Founders

Type	Count	Percent (%)
Female	1953	7.4
White	21379	81.0
East Asian	1600	6.1
South Asian	2395	9.1
Hispanic	870	3.3
Black	76	0.3

Table 16: Venture-Backed Founder Matching (Venture Source)

In the regression table, each observation is a founder-founder pair, where each founder is matched to all pseudo-founders in the same state-year-industry. The independent variable Same Ethnicity is a binary indicator. We include start-ups with at least two founders. Statistical significance at the 1%, 5%, and 10% levels are denoted by ***, **, and *, respectively.

	Dependent Variable: Real Match	
	(1)	(2)
Same Ethnicity	0.0086***	
	(0.000)	
Both East Asian		0.042***
		(0.002)
Both South Asian		0.041***
		(0.002)
Both Hispanic		0.030***
		(0.005)
Both Black		0.24**
		(0.102)
Both White		0.0066***
		(0.000)
Both Female	0.023***	0.023***
	(0.003)	(0.003)
Both Male	0.0017***	0.0018***
	(0.001)	(0.001)
Founder Count	0.010***	0.010***
	(0.000)	(0.000)
Observations	1469730	1469730
R-Squared	0.142	0.144
Year FE	Y	Y
Industry FE	Y	Y
State FE	Y	Y

Table 17: Correlation between Start-up Team Diversity and Performance (Venture Source)

The table reports the performance results. The independent variable Race/Ethnicity Score is constructed in the same manner as our main regression results. We include start-ups with at least two founders. Statistical significance at the 1%, 5%, and 10% levels are denoted by ***, **, and *, respectively.

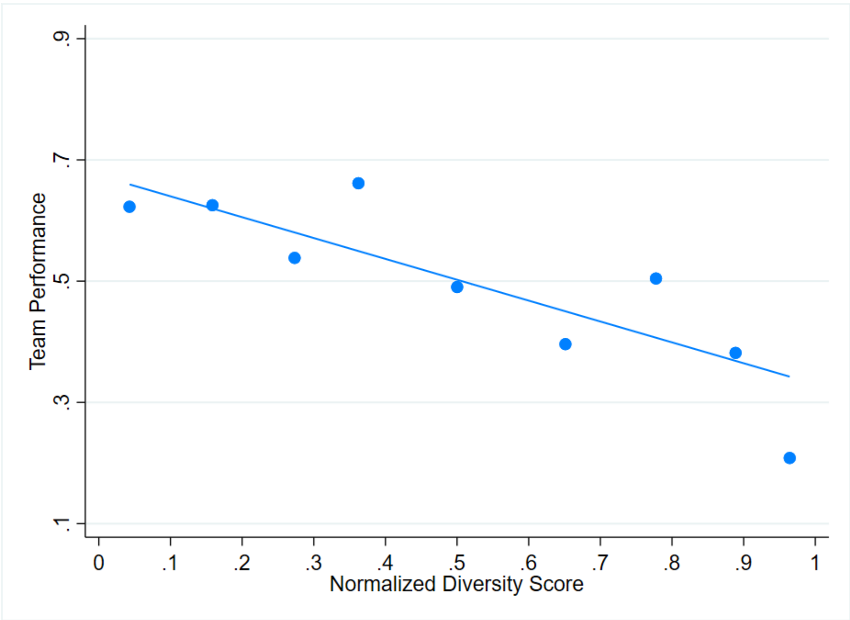
	Dependent Variable: IPO	
	(1)	(2)
Race/Ethnicity Score	-0.011	-0.011
	(0.007)	(0.007)
Founder Count	0.020***	0.020***
	(0.003)	(0.003)
Number of Female Founders		-0.0051
		(0.006)
Number of Top Schools		0.0040
		(0.005)
Constant	0.42***	0.43***
	(0.090)	(0.090)
Observations	8156	8156
R-Squared	0.145	0.145
Year FE	Y	Y
Industry FE	Y	Y
State FE	Y	Y

A Appendix Figures

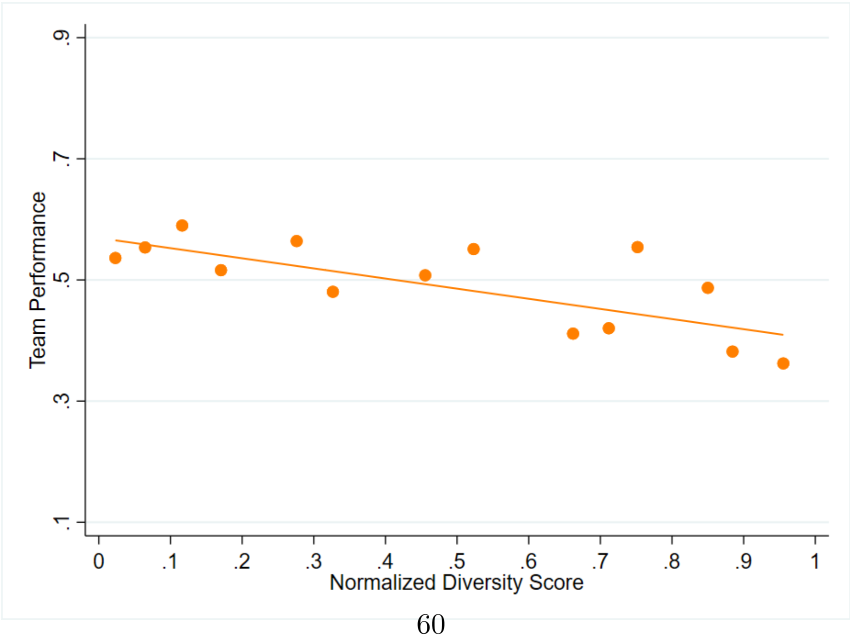
Figure A.1: Horizontal Team Diversity and Team Performance

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 normalized race/ethnicity diversity score of the team, measures as the percentile of the fraction of mismatched race or ethnicity ties within a team against a distribution of randomly formed teams. Lower scores represent more homogeneous teams, whereas higher scores represent more diverse teams.

(a) Diversity vs. Performance with Random Team Assignment (2013)



(b) Diversity vs. Performance with Voluntary Team Formation (2014-2016)



B Appendix Tables

Table B.1: 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	2394	64.98%
2	India	172	4.67%
3	Canada	125	3.39%
4	China	76	2.06%
5	United Kingdom	59	1.60%
6	Brazil	52	1.41%
7	Australia	46	1.25%
8	Germany	45	1.22%
9	France	45	1.22%
10	Israel	33	0.90%
11	South Korea	30	0.81%
12	Japan	28	0.76%
13	Turkey	28	0.76%
14	Mexico	28	0.76%
15	Argentina	27	0.73%
16	Russia	25	0.68%
17	Lebanon	25	0.68%
18	Spain	24	0.65%
19	Nigeria	23	0.62%
20	Chile	19	0.52%
	Total	3684	100.00%

Table B.2: Summary Statistics on Career Interests

This table reports summary statistics on the ten most popular career interests listed by HBS students on their Class Card profiles. Entrepreneurship/start-up was only made available as a choice only for class of 2014 and later.

	Career Interests	Percent
1	Finance	47.6%
2	Technology	33.9%
3	Consulting	25.8%
4	Entrepreneurship/Startup	24.3%
5	Consumer Products	19.1%
6	Leisure	16.9%
7	Healthcare	14.3%
8	Community Development	12.4%
9	Education	11.6%
10	Government	11.2%
	Total number of students	3684
	Median number of career interests	3
	Fraction of students listing at least 1 career interest	80.9%

Table B.3: Summary Statistics on Extracurricular Activities

This table reports summary statistics on the most popular extracurricular activities listed by HBS students on their Class Card profiles.

	Extracurricular Clubs	Percent
1	VCPE Club	25.7%
2	Entrepreneurship Club	18.9%
3	Socialenterprise Club	15.5%
4	Healthcare Club	15.0%
5	Techmedia Club	12.0%
6	Retail Club	10.8%
7	Wine Club	10.6%
8	Investment Club	10.5%
9	General Management Club	10.0%
10	Women Club	9.6%
	Number of students listing at least 1 club	1102
	Median number of clubs	3
	Fraction of students listing at least 1 club	29.9%

Table B.4: 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) Finance	(2) Technology	(3) Consulting	(4) Healthcare	(5) Retail	(6) Top School	(7) Start-up Exp
Female	-0.020	-0.0043	0.094***	0.0000053	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)
Black	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	3684	3684	3684	3684	3684	3684	3684
R-Squared	0.008	0.008	0.039	0.015	0.002	0.121	0.007
Year FE	Y	Y	Y	Y	Y	Y	Y

Table B.5: Correlation between Diversity Score and Individual Quality

This table regresses various measures of individual quality on race/ethnicity diversity scores. Each observation is a team.

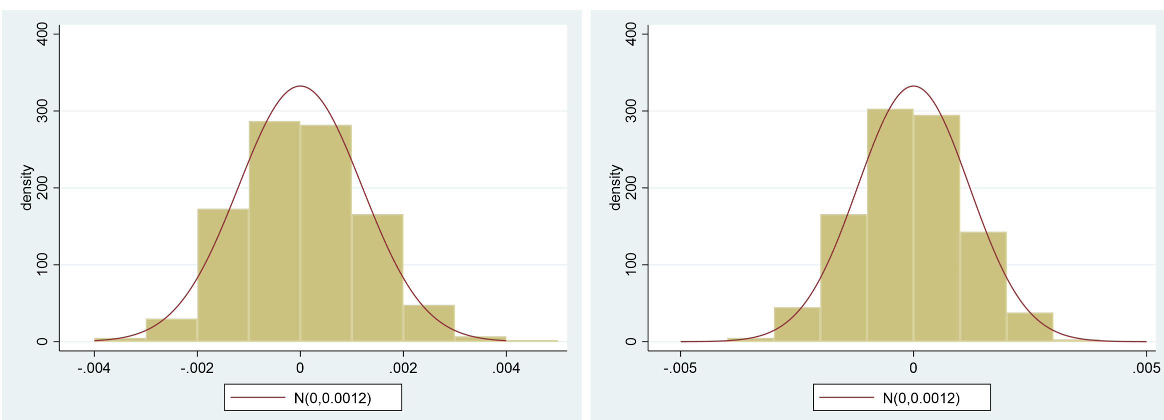
	Dependent Variable: Performance					
	Randomly Assigned (2013)			Voluntarily Formed (2014-2016)		
	(1) Top School Ratio	(2) Start-up Ratio	(3) Honors Ratio	(4) Top School Ratio	(5) Start-up Ratio	(6) Honors Ratio
Race/Ethnicity Score	-0.048	-0.059	-0.0040	-0.17***	-0.011	-0.14***
	(0.114)	(0.036)	(0.060)	(0.060)	(0.020)	(0.023)
Team Member Count	-0.057	0.010	0.041	0.046**	0.015*	0.031*
	(0.068)	(0.018)	(0.038)	(0.020)	(0.008)	(0.016)
Constant	0.82*	0.011	-0.12	0.24*	-0.054	0.048
	(0.401)	(0.108)	(0.263)	(0.138)	(0.054)	(0.094)
Observations	150	150	150	480	480	480
Year FE	N/A	N/A	N/A	N/A	N/A	N/A

Table B.6: Randomization Test of Homophily in Team Formation

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 that the simulated value is greater than the actual matching coefficient.

	Subsample: 2014-2016				
	Simulated Matching Coefficients		Actual Matching 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%

(a) Distribution of Race/Ethnicity Tie Coefficients (Left) and Gender Tie Coefficients (Right)



(b) Distribution of Race/Ethnicity Tie Coefficients (Left) and Gender Tie Coefficients (Right)

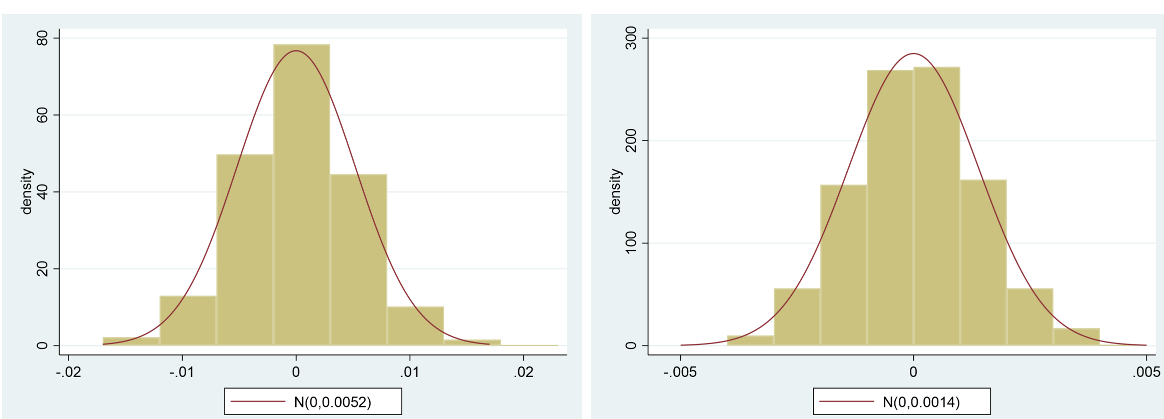


Table B.7: Detailed Matching Regression: International Students

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 1 if the students are teammates. The independent variables are ethnicity characteristics, equaling 1 if both students are from the same region. Robust standard errors are clustered at the student level.

	Dependent Variable: Real Match	
	(1) Voluntarily Formed Teams (2014-2016)	(2) 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	254318	81368
R-Squared	0.002	0.000
Year FE	Y	Y

Table B.8: Controlling for All Confounding Characteristics

This table regresses team performance on team diversity scores controlling for a variety of student characteristics. 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 error is clustered at year-section level. Statistical significance at the 1%, 5%, and 10% levels are denoted by ***, **, and *, respectively.

	Dependent Variable: Performance			
	Randomly Assigned (2013)		Voluntarily Formed (2014-2016)	
	(1)	(2)	(3)	(4)
Race/Ethnicity Score	-0.48***	-0.48	-0.18***	-0.050
	(0.110)	(0.303)	(0.054)	(0.096)
Finance %		-0.056		-0.046
		(0.167)		(0.065)
Consulting %		0.017		-0.080
		(0.187)		(0.088)
Technology %		0.13		0.25**
		(0.126)		(0.109)
Start-up Ratio		0.43		0.24
		(0.432)		(0.148)
Top School Ratio		0.11		-0.013
		(0.062)		(0.059)
Honor Student Ratio		0.32		0.29***
		(0.180)		(0.095)
English Speaker %		-0.045		0.19
		(0.315)		(0.133)
% Female		-0.077		0.16***
		(0.337)		(0.055)
% Black		-0.37		-0.12
		(0.323)		(0.170)
% Asian		-0.0087		-0.21*
		(0.220)		(0.117)
% Latino		-0.21		0.17
		(0.310)		(0.150)
% White		-0.090		-0.042
		(0.194)		(0.138)
Team Member Count	0.047	0.072	0.088***	0.078***
	(0.059)	(0.070)	(0.026)	(0.027)
Constant	0.59	0.46	0.045	-0.17
	(0.358)	(0.665)	(0.165)	(0.205)
Observations	150	150	480	480
R-Squared	0.087	0.157	0.049	0.132
Year FE	N/A	N/A	Y	Y

C Appendix Model

Lemma 3.1. *(Optimal effort is decreasing in diversity.) Under mild regularity conditions, namely,*

(i) Overall utility $U(g, e)$ is concave in effort e

(ii) The cost of effort is increasing in team discord $\frac{\partial^2 C}{\partial d \partial e} > 0$

(iii) Quality and effort enters separately into producing performance $\frac{\partial^2 F}{\partial q \partial e} = 0$

The optimal level of effort $e^*(g)$ is decreasing in group diversity

$$\frac{\mathrm{d}e^*}{\mathrm{d}g} < 0. \tag{C.1}$$

Proof. For a given group composition g, the team makes a choice for the optimal level of effort $e^*(g)$ which solves the first order condition

$$\frac{\partial F}{\partial e} - \frac{\partial C}{\partial e} = 0, \tag{C.2}$$

where the marginal gain from the increased performance from more effort equals to the marginal cost of exerting more effort.

Next, consider a composition change that increases diversity. We take the total derivative of Equation (C.2) with respect to g as follows

$$\frac{\mathrm{d}}{\mathrm{d}g}\left(\frac{\partial F(q(g), e^*(g))}{\partial e} - \frac{\partial C(e^*(g), d(g))}{\partial e}\right) = 0$$

which yields

$$\frac{\mathrm{d}e^*}{\mathrm{d}g} = -\left(\underbrace{\frac{\partial^2 F}{\partial q \partial e}\frac{\mathrm{d}q}{\mathrm{d}g}}_{=0} - \underbrace{\frac{\partial^2 C}{\partial d \partial e}\frac{\mathrm{d}d}{\mathrm{d}g}}_{>0}\right) \Big/ \underbrace{\left(\frac{\partial^2 F}{\partial e^2} - \frac{\partial^2 C}{\partial e^2}\right)}_{<0 \text{ by concavity of } U} < 0. \tag{C.3}$$

□

It is worth noting that the assumption that quality and effort enters separately into performance is a sufficient but not necessary condition for the optimal effort to decrease in diversity. In particular, the second term of $\frac{\partial^2 F}{\partial q \partial e}\frac{\mathrm{d}q}{\mathrm{d}g}$, the unconditional relationship between quality and diversity could well be close to zero, and it would also generate the desired comparative statistic. Moreover, empirically, it is true that there is no statistically significant relationship between group diversity and individual quality, approximated by the fraction of students who earn honors, as shown in Appendix Table B.5. Therefore, in the case where $\frac{\mathrm{d}q}{\mathrm{d}g} \approx 0$, optimal effort is still decreasing in diversity, regardless whether quality and effort can enter as substitutes or complements into the production of performance.

Lemma 3.4. *(Positive correlation between observed and unobserved diversity.) For g^o and g^u that both positively enter into the cost of effort $d(g^o, g^u)$, they are positively correlated when teams are exogenously assigned*

$$corr(g^u, g^o)_{g \sim \mathcal{F}^X} > 0. \tag{C.4}$$

Proof. To simplify exposition, consider two individual binary characteristics $X = (X^o, X^u)$ that are positively correlated

$$X^o \sim \text{Bernoulli}(p_o) \tag{C.5}$$
$$X^u \sim \text{Bernoulli}(p_u) \tag{C.6}$$

Moreover, to allow for positive correlation between X^o and X^u, without loss of generality, we can allow X^u to be a mixing distribution based on X^o

$$X^u = \begin{cases} \text{Bernoulli}(p_u + (1 - p_o)\theta), & \text{if } X^o = 1 \\ \text{Bernoulli}(p_u - p_o\theta), & \text{if } X^o = 0 \end{cases} \tag{C.7}$$

where $\theta > 0$. As such, the correlation becomes

$$corr(X^o, X^u) = \mathbb{E}[X^o X^u] - \mathbb{E}[X^o]\mathbb{E}[X^u] = p_o(1 - p_o)\theta > 0. \tag{C.8}$$

Then, let M^o and M^u denote whether two randomly drawn individuals X_1 and X_2 are matched in their observed and unobserved characteristics

$$M^o = \mathbb{1}\{X_1^o = X_2^o\} \tag{C.9}$$
$$M^u = \mathbb{1}\{X_1^u = X_2^u\}. \tag{C.10}$$

The objective is to see whether $corr(M^o, M^u)$ is also positive.

We can re-write components of the correlation as follows.

$$\mathbb{E}[M^o] = \mathbb{E}[\mathbb{1}\{X_1^o = X_2^o\}] \tag{C.11}$$
$$= \mathbb{E}[\mathbb{1}\{X_1^o = 1, X_2^o = 1\}] + \mathbb{E}[\mathbb{1}\{X_1^o = 0, X_2^o = 0\}] \tag{C.12}$$
$$= \mathbb{E}[\mathbb{1}\{X_1^o = 1\}]\mathbb{E}[\mathbb{1}\{X_2^o = 1\}] + \mathbb{E}[\mathbb{1}\{X_1^o = 0\}]\mathbb{E}[\mathbb{1}\{X_2^o = 0\}] \tag{C.13}$$
$$= \mathbb{E}[X_1^o]\mathbb{E}[X_2^o] + \mathbb{E}[1 - X_1^o]\mathbb{E}[1 - X_2^o] \tag{C.14}$$
$$= p_o^2 + (1 - p_o)^2 \tag{C.15}$$

where we take advantage of the fact that X_1 and X_2 are drawn independently. Similarly, we have

$$\mathbb{E}[M^u] = p_u^2 + (1 - p_u)^2. \tag{C.16}$$

Next, consider the interaction term

$$\mathbb{E}[M^o M^u] = \mathbb{E}[\mathbb{1}\{X_1^o = X_2^o\}\mathbb{1}\{X_1^u = X_2^u\}] \tag{C.17}$$
$$= \mathbb{E}[\mathbb{1}\{X_1^o = 1, X_2^o = 1, X_1^u = 1, X_2^u = 1\}] \tag{C.18}$$
$$+ \mathbb{E}[\mathbb{1}\{X_1^o = 1, X_2^o = 1, X_1^u = 0, X_2^u = 0\}] \tag{C.19}$$
$$+ \mathbb{E}[\mathbb{1}\{X_1^o = 0, X_2^o = 0, X_1^u = 1, X_2^u = 1\}] \tag{C.20}$$
$$+ \mathbb{E}[\mathbb{1}\{X_1^o = 0, X_2^o = 0, X_1^u = 0, X_2^u = 0\}] \tag{C.21}$$
$$= \mathbb{E}[X^o X^u]^2 + \mathbb{E}[X^o(1 - X^u)]^2 + \mathbb{E}[X^u(1 - X^o)]^2 + \mathbb{E}[(1 - X^o)(1 - X^u)]^2 \tag{C.22}$$
$$= (p_o(p_u + (1 - p_o)\theta))^2 + (p_o(1 - p_u - (1 - p_o)\theta))^2 \tag{C.23}$$
$$+ ((1 - p_o)(p_u + (1 - p_o)\theta))^2 + ((1 - p_o)(1 - p_u - (1 - p_o)\theta))^2 \tag{C.24}$$

Putting all terms together, we have

$$corr(M^o, M^u) = \mathbb{E}[M^o M^u] - \mathbb{E}[M^o]\mathbb{E}[M^u] \tag{C.25}$$
$$= 2\theta((1 - 2p_o)(1 - 2p_u) + 2\theta) \tag{C.26}$$
$$> 2\theta(1 - 2p_o)(1 - 2p_u) \tag{C.27}$$
$$\geq 0. \tag{C.28}$$

Note that the last step technically requires $p_o \leq 1/2$ and $p_u \leq 1/2$, but it is not an issue because we can always redefine the binary characteristics so that they take on a value of one for the minority outcome.

Lastly, since diversity in characteristics is just the opposite of alignment in characteristics, we have

$$corr(G^o, G^u) = \mathbb{E}[G^o G^u] - \mathbb{E}[G^o]\mathbb{E}[G^u] \tag{C.29}$$
$$= \mathbb{E}[(1 - M^o)(1 - M^u)] - \mathbb{E}[1 - M^o]\mathbb{E}[1 - M^u] \tag{C.30}$$
$$= (1 - \mathbb{E}[M^o] - \mathbb{E}[M^u] + \mathbb{E}[M^o M^u]) - (1 - \mathbb{E}[M^o] - \mathbb{E}[M^u] + \mathbb{E}[M^o]\mathbb{E}[M^u]) \tag{C.31}$$
$$= corr(M^o, M^u) > 0. \tag{C.32}$$

□

Lemma 3.5. *(Selection weakens the correlation between observed and unobserved diversity) The correlation between observed and unobserved measures of diversity becomes weakened when teams are endogenously formed:*

$$corr(g^u, g^o)_{g \sim \mathcal{F}^X} > corr(g^u, g^o)_{g \sim \mathcal{F}^N}. \tag{C.33}$$

Proof. Consider the same set up as the previous lemma. However, when selection is allowed, X_1 and X_2 are no longer independently drawn. In particular, we allow for different marginal distribution of (X_1^u, X_2^u) depending on the realization of (X_1^o, X_2^o). Concretely, we consider the distribution of X^u to be a mixing of two latent types X^L, high and low, with equal probability

$$X^u = \begin{cases} \text{Bernoulli}(p_u + \eta), & \text{if } X^L = 1 \\ \text{Bernoulli}(p_u - \eta), & \text{if } X^L = 0 \end{cases} \tag{C.34}$$

Then, when there is a mismatch on the observed characteristics $X_1^o = X_2^o$, we allow individuals to match on their latent types $X_1^L = X_2^L$. Conversely, when there is a match on the observed characteristics, individuals will be mismatched on their latent types $X_1^L \neq X_2^L$.

In the absence of a full model for student matching, the latent type is introduced as a modeling device to capture the statistical property where student pairs that are mismatched on the observables are more likely to select on their unobservables. The sign of parameter η governs whether they are selecting to be more ($\eta > 0$) or less ($\eta < 0$) matched on the unobservables than randomly draws ($\eta = 0$). In other words, the probability of matching $X_1^u = X_2^u$ increases with $\eta > 0$, and the probability of mismatching $X_1^u \neq X_2^u$ with $\eta < 0$.[24]

Moreover, the introduction of the latent type preserves the unconditional correlation between X^o and X^u in the population in the following sense

$$X^u = \begin{cases} \text{Bernoulli}(p_u + (1 - p_o)\theta + \eta), & \text{if } X^L = 1, X^o = 1 \\ \text{Bernoulli}(p_u - p_o)\theta + \eta), & \text{if } X^L = 1, X^o = 0 \\ \text{Bernoulli}(p_u + (1 - p_o)\theta - \eta), & \text{if } X^L = 0, X^o = 1 \\ \text{Bernoulli}(p_u - p_o)\theta - \eta), & \text{if } X^L = 0, X^o = 0 \end{cases} \tag{C.35}$$

[24]A regularity condition is that η moderately sized so as to $p_u + \eta < 1/2$. However, this is not an issue because there always exists a value η that achieves a certain level of matching probability without violating the condition because of symmetry around 1/2.

Thus, we can compute the correlation between M^o and M^u analogously

$$\begin{aligned}
\mathbb{E}[M^oM^u]_{\mathcal{F}^{\mathcal{N}}} =& \mathbb{E}[\mathbb{1}\{X_1^o = X_2^o\}\,\mathbb{1}\{X_1^u = X_2^u\}] && \text{(C.36)}\\
=& \mathbb{E}[\mathbb{1}\{X_1^o = 1, X_2^o = 1, X_1^u = 1, X_2^u = 1\}] && \text{(C.37)}\\
&+ \mathbb{E}[\mathbb{1}\{X_1^o = 1, X_2^o = 1, X_1^u = 0, X_2^u = 0\}] && \text{(C.38)}\\
&+ \mathbb{E}[\mathbb{1}\{X_1^o = 0, X_2^o = 0, X_1^u = 1, X_2^u = 1\}] && \text{(C.39)}\\
&+ \mathbb{E}[\mathbb{1}\{X_1^o = 0, X_2^o = 0, X_1^u = 0, X_2^u = 0\}] && \text{(C.40)}\\
=& \mathbb{E}\left[\mathbb{1}\left\{X_1^o = 1, X_2^o = 1, X_1^u = 1, X_2^u = 1, X_1^L = 1, X_2^L = 0\right\}\right] && \text{(C.41)}\\
&+ \mathbb{E}\left[\mathbb{1}\left\{X_1^o = 1, X_2^o = 1, X_1^u = 1, X_2^u = 1, X_1^L = 0, X_2^L = 1\right\}\right] && \text{(C.42)}\\
&+ \mathbb{E}\left[\mathbb{1}\left\{X_1^o = 1, X_2^o = 1, X_1^u = 0, X_2^u = 0, X_1^L = 1, X_2^L = 0\right\}\right] && \text{(C.43)}\\
&+ \mathbb{E}\left[\mathbb{1}\left\{X_1^o = 1, X_2^o = 1, X_1^u = 0, X_2^u = 0, X_1^L = 0, X_2^L = 1\right\}\right] && \text{(C.44)}\\
&+ \mathbb{E}\left[\mathbb{1}\left\{X_1^o = 0, X_2^o = 0, X_1^u = 0, X_2^u = 0, X_1^L = 1, X_2^L = 0\right\}\right] && \text{(C.45)}\\
&+ \mathbb{E}\left[\mathbb{1}\left\{X_1^o = 0, X_2^o = 0, X_1^u = 0, X_2^u = 0, X_1^L = 0, X_2^L = 1\right\}\right] && \text{(C.46)}\\
&+ \mathbb{E}\left[\mathbb{1}\left\{X_1^o = 0, X_2^o = 0, X_1^u = 1, X_2^u = 1, X_1^L = 1, X_2^L = 0\right\}\right] && \text{(C.47)}\\
&+ \mathbb{E}\left[\mathbb{1}\left\{X_1^o = 0, X_2^o = 0, X_1^u = 1, X_2^u = 1, X_1^L = 0, X_2^L = 1\right\}\right] && \text{(C.48)}\\
=& p_o^2(p_u + (1-p_o)\theta + \eta)(p_u + (1-p_o)\theta + \eta) && \text{(C.49)}\\
&+ p_o^2(1 - p_u - (1-p_o)\theta + \eta)(1 - p_u - (1-p_o)\theta + \eta) && \text{(C.50)}\\
&+ (1-p_o)^2(p_u + (1-p_o)\theta + \eta)(p_u + (1-p_o)\theta + \eta) && \text{(C.51)}\\
&+ (1-p_o)^2(1 - p_u - (1-p_o)\theta + \eta)(1 - p_u - (1-p_o)\theta + \eta) && \text{(C.52)}
\end{aligned}$$

Putting the two together, the difference in the correlation in the match characteristics across different team assignment mechanisms arises from the difference in the interaction term:

$$\begin{aligned}
corr(G^o, G^u)_{\mathcal{F}^{\mathcal{X}}} - corr(G^o, G^u)_{\mathcal{F}^{\mathcal{N}}} &= corr(M^o, M^u)_{\mathcal{F}^{\mathcal{X}}} - corr(M^o, M^u)_{\mathcal{F}^{\mathcal{N}}} && \text{(C.53)}\\
&= \mathbb{E}[M^oM^u]_{\mathcal{F}^{\mathcal{X}}} - \mathbb{E}[M^oM^u]_{\mathcal{F}^{\mathcal{N}}} && \text{(C.54)}\\
&= \frac{1}{2}\eta^2(1 - 2(1-p_o)p_o) && \text{(C.55)}\\
&> 0. && \text{(C.56)}
\end{aligned}$$

Note that the last term is positive because $(1-p_o)p_o$ has a max value of $1/2$ for $x \in [0,1]$. More importantly, we note that the correlation weakens with endogenous selection regardless whether they select more or less heavily on unobservables than the exogenously assignment, i.e., regardless whether η is positive or negative because it enters as a squared term.

$\square$

Proposition 3.6. *(Match on Observables) If the quality channel of diversity is dominated by the direct cost of diversity, namely,* $\frac{\partial F}{\partial q}\frac{\mathrm{d}q}{\mathrm{d}g} - \frac{\partial C}{\partial d}\frac{\mathrm{d}d}{\mathrm{d}g} \leq 0$*, then,* $g^o \sim \mathcal{F}^N$ *will be first-order stochastically dominated by* $g^o \sim \mathcal{F}^X$*,*

$$\forall \tilde{g}: \ \mathbb{P}_{\mathcal{F}^X}\{g^o < \tilde{g}\} \leq \mathbb{P}_{\mathcal{F}^N}\{g^o < \tilde{g}\}. \tag{C.57}$$

Proof. Consider the baseline model of utility with optimal effort $U(g) = U(g, e^*(g))$ with $g \sim \mathcal{F}^X$ under exogenous assignment:

$$\frac{\mathrm{d}U(g)}{\mathrm{d}g} = \frac{\mathrm{d}}{\mathrm{d}g}F(q(g), e^*(g)) - C(e^*(g), d(g)) \tag{C.58}$$

$$= \frac{\partial F}{\partial q}\frac{\mathrm{d}q}{\mathrm{d}g} + \frac{\partial F}{\partial e}\frac{\mathrm{d}e^*}{\mathrm{d}g} - \frac{\partial C}{\partial e}\frac{\mathrm{d}e^*}{\mathrm{d}g} - \frac{\partial C}{\partial d}\frac{\mathrm{d}d}{\mathrm{d}g} \tag{C.59}$$

$$= \underbrace{\frac{\partial F}{\partial q}\frac{\mathrm{d}q}{\mathrm{d}g}}_{\substack{\text{The Quality Channel}\\ \text{of Diversity}}} - \underbrace{\frac{\partial C}{\partial d}\frac{\mathrm{d}d}{\mathrm{d}g}}_{\substack{\text{The Direct Cost}\\ \text{of Diversity}}} \tag{C.60}$$

where the second to last step uses the Envelope Theorem on the optimality of $e^*(g)$ such that $\frac{\partial F}{\partial e} = \frac{\partial C}{\partial e}$. Consequently, if $\frac{\mathrm{d}U(g)}{\mathrm{d}g} < 0$, i.e., the direct utility cost of diversity dominates the quality channel of diversity, it is utility improving to choose a different teammate to decrease diversity. Thus, endogenously-formed teams will be less diverse in a distributional sense than exogenously-formed teams. Since the researcher can only observe g^o, assuming that observed and unobserved characteristics are positive correlated in the population, we expect that the observed component of diversity to also decrease. $\square$

Proposition 3.7. *(Match on Unobservables) If the quality channel of unobserved diversity is dominated by the direct cost channel, namely,* $\frac{\partial F}{\partial q}\frac{\partial q}{\partial g^u} - \frac{\partial C}{\partial d}\frac{\partial d}{\partial g^u} \leq 0$*, then endogenously formed teams are more likely to match on unobservable characteristics than exogenously assigned teams, conditioning on the same level of observed diversity,*

$$\mathbb{E}_{\mathcal{F}^X}[g^u | g^o] \geq \mathbb{E}_{\mathcal{F}^N}[g^u | g^o]. \tag{C.61}$$

Proof. Consider the partial derivative of utility with respect to g^u holding g^o fixed, $(g^o, g^u) \sim \mathcal{F}^X$

$$\frac{\partial U(g^o, g^u)}{\partial g^u} = \frac{\partial U(g^o, g^u, e^*(g^o, g^y))}{\partial g^u} \tag{C.62}$$

$$= \frac{\partial F}{\partial q}\frac{\partial q}{\partial g^u} + \frac{\partial F}{\partial e}\frac{\partial e^*}{\partial g^u} - \frac{\partial C}{\partial e}\frac{\partial e^*}{\partial g^u} - \frac{\partial C}{\partial d}\frac{\partial d}{\partial g^u} \tag{C.63}$$

$$= \underbrace{\frac{\partial F}{\partial q}\frac{\partial q}{\partial g^u}}_{\substack{\text{The Quality Channel}\\ \text{of Unobserved Diversity}}} - \underbrace{\frac{\partial C}{\partial d}\frac{\partial d}{\partial g^u}}_{\substack{\text{The Direct Cost}\\ \text{of Unobserved Diversity}}} \tag{C.64}$$

where the second to last step also uses the Envelope Theorem. Consequently, if $\frac{\partial U}{\partial g^u} < 0$, it is utility improving to choose a different teammate to decrease unobserved diversity, condition on observed diversity g^o. Thus, the expected level of unobserved diversity among endogenously-formed teams will be lower than exogenously-assigned teams.

$\square$

Note that, in principle, we can also make a statement about the stochastic dominance of the distribution of $g^u|g^o_{\mathcal{FX}}$ over $g^u|g^o_{\mathcal{FN}}$, but we opt for a model prediction that based on the comparison of means, because it is difficult to estimate the conditional distributions precisely.