Team Diversity and Individual Performance

Swarnodeep Homroy[a] and Kwok Tong Soo[b]

[a]University of Groningen; [b]Lancaster University

Abstract

Using a field experiment, we investigate the impact of working in a diverse team on team and individual performance. We find no short-term effect of team diversity on team performance, but a positive (albeit small) effect on subsequent individual performance. In the latter case it is diversity in terms of nationality which matters, not diversity in terms of gender or ability. Our results suggest that the gains from asymmetries in work teams are assimilated by individuals.

Keywords: Team composition; diversity; individual performance; team performance; field experiment

Word Count: 7147

1 INTRODUCTION

This paper is about diversity in team composition, and the effect diversity may have on the *ex post* performance of individual group members. Even if members of a group are *ex ante* identical to each other in terms of their abilities, Smith (1776) shows that a group of workers will be able to increase their total output through specialization. If, in addition, workers have diverse abilities, then specialization according to each worker's individual skills would enhance the group performance (Cox, Lobel and McLeod, 1991; Watson, Kumar and Michaelsen, 1993).[1] On the other hand, it can be argued that coordination costs in diverse teams reduce collective performance (O'Reilly, Caldwell and Barnett, 1989; Zenger and Lawrence, 1989). A more diverse group is typically more difficult to coordinate than a less diverse one. The potential gains and costs of diversity is an important issue for resource allocation as the labour force becomes more diverse due to the increase in female participation and international migration (Tomasakovic-Davy et al. 2006; Ely, Padavic and Thomas, 2012). For a firm to be willing to employ a diverse workforce, the gains from having diverse work teams must outweigh the potential coordination costs.

A number of studies across personnel economics, sports economics, and corporate governance have focused on the impact of team diversity on collective performance and the results are inconclusive (Kochan et al. 2003; Zarraga and Bonache, 2005). Diversity is a multi-dimensional concept, and many of these papers focus on a specific measure of diversity. For example, Lazear (1999) argues that if its members have a diverse range of cultures, a team benefits from greater collective knowledge and skills. Lee (2014), Nathan (2013, 2015), and Nathan and Lee (2013) present a range of evidence on the generally positive impact of ethnic and gender diversity on performance and innovation in UK firms and cities. In sports, Papps et al (2011) find that there is an optimal degree of dispersion in player ability in a major

[1] Organizational studies use slightly different definitions for "group" and "team". However, a vast majority of the economics literature uses these terms interchangeably. We follow this latter practice. The fundamental notion is that of a collection of individuals working together towards a common goal within a given timeframe.

league baseball team. Teams with a balanced spread of ability tend to do better than teams in which dispersion in individual abilities is too high and too low. Kahane et al (2013) find that National Hockey League teams that employ a higher proportion of European players performed better. Similarly, research in corporate governance has shown that there is little impact of boardroom diversity on corporate performance (Adams and Ferreira, 2009; Green and Homroy, 2018).

Our focus in this paper is slightly different from the preceding literature. First, whilst we investigate the effect of team diversity on group performance, we also explore its effects on subsequent individual performance, both in the short-run and long-run. The literature on peer effects on individual performance is small but growing (Bandiera et al, 2010; Hoogendoorn et al, 2013). Second, we explore the many dimensions of diversity, to try and understand which types of diversity have more important effects on performance.

Identifying the causal impact of team diversity on performance is fraught with econometric issues like assortative sorting, reverse causality and unobserved heterogeneity. To address these concerns, we use a field experiment from a course at Lancaster University which has a compulsory assessed team project, followed by an individual assessment. The formation of the project team is randomised at the beginning of the academic year. This mitigates a major concern in the peer-effect literature related to endogenous team formation. This approach has some benefits over studies based on observational data and laboratory settings. For instance, Becker (1973) shows theoretically that output-maximising partnerships involve positive assortative matching when traits are complements. To circumvent this problem of self-selection, Katz et al. (2001), Sacerdote (2001), Falk and Ichino (2006) and Carrell et al. (2015) use (field and laboratory) experimental settings to randomly assign individuals to different peer groups. These papers all find clean evidence of peer effects. Vogel et al (2014) use a field experiment to show that team diversity affects funding decisions of entrepreneurial teams.

Relative to non-experimental settings, our field experiment approach helps to avoid the problem of assortative selection into groups (Duflo et al, 2008; Hoogendoorn, Oosterbeek and van Praag, 2013). A major advantage of our set-up is the sequential timing of the group and the individual activities, which allows us to focus on the direction of causality. In addition, there is no sample attrition problem in our setting. Unlike the experimental-laboratory approach such as Falk and Ichino (2006), our analysis is based on an actual, assessed, non-experimental task which is spread over several weeks and is not restricted to a few laboratory sessions. The institutional setting is largely replicable in a wide range of organizations, addressing concerns about the external validity of such experiments. The usual critique of the lack of incentives in experimental settings also doesn't apply because the students are actually assessed on the tasks given. Therefore, there are real consequences to the individual of failure of the task. Moreover, because the group project is part of the course design, we have data for several cohorts of students who are engaged in the same tasks. Even so, this is not a conventional randomized control trial (RCT), and not all contaminating factors can be explicitly controlled for. In subsequent sections we discuss the field experiment, the potential contaminating factors, and how we attempt to address these through the experimental design.

We distinguish between three types of diversity: nationality, gender, and ability. Our first main result is that group diversity of any of the three types has no statistically significant effect on group performance. This suggests that the gains from diverse skill sets may be counter-balanced by higher costs of coordination. However, working in nationally-diverse teams (but not gender-diverse or ability-diverse teams) enhances *ex-post* individual performance of the team members. The effect is small, but statistically significant. The performance gain at the individual level persists beyond the end of the experiment. As we discuss later, there are potentially two counter-balancing economic mechanisms at work: first, individuals' ability to work effectively in a heterogeneous team, and second, learning from being in an environment with diverse team members. Our results suggest that the second effect

is more important than the first. In addition, in sensitivity analysis, a broader definition of sociocultural diversity has a positive and significant effect on both group and individual performance. This may indicate that our measure of national diversity may not be able to fully capture how diversity can affect performance.

A second focus of this paper is to take a long term approach to measuring group and individual performance. Many previous studies have focused on concurrent group performance (Chatman and Flynn, 2001; Polzer, Milton and Swann, 2002). One possible mechanism of the impact of team diversity is through gains in the skills and productivity of individuals. It is important to understand the performance impact of working in a team on individuals because modern firms are organised around teams which are increasingly diverse. One reason for this is that team decisions are more likely to comply with economic rationality than individual decisions (Rockenbach, Sadrieh and Mathauschek, 2007). If workers participate in repeated team interactions, the gains in individual productivity from working in a diverse team will enhance the performance of future work teams and increase average labour productivity.

We investigate whether or not this gain exhibits systematic patterns across different groups of students. There is no differential gain in productivity either by nationality or ability. This contrasts with Bandiera et al (2010) who show that individuals who work with more productive peers become more productive themselves, and *vice versa*. Therefore, there seems to be a performance spillover from diversity for all members of a diverse work team. Even though diverse groups may not collectively perform any better than homogeneous groups, exposure to diverse work groups enhances the subsequent individual performance of team members.

Our paper is closely related to Hoogendoorn et al. (2013) in using a field experiment to estimate the effect of group diversity on performance. Whilst Hoogendoorn et al. (2013) examines the effect of gender diversity on contemporaneous group and individual performance

of students, we use both gender and national diversity in groups and examine *ex-post* and longer term individual performance. Our paper complements the findings of Hoogendoorn et al. (2013) in suggesting that diversity of nationality of group members affects longer term individual performance, even if the economic effect on the immediate group performance is small.

Section 2 discusses the institutional background, and the experimental set up, section 3 presents the results, and section 4 concludes.

2 EXPERIMENTAL SETUP

2.1 Institutional Setting

The central question is how team diversity impacts upon collective, and subsequent individual performance. To investigate this, we use data from a first year, compulsory, full-year module on Quantitative Methods for Economics (henceforth Quantitative Methods) at Lancaster University. This module is compulsory for all students on the single-major BSc in Economics, and is an optional module for joint majors and students on the BSc in Business Economics. Students study three full-year modules in their first year.[2]

The choice of this module is driven by the presence of a teamwork component of assessment, followed by an individual piece of assessment. Quantitative Methods is a 25-week course comprising 12 weeks each of basic mathematics and statistics with one week set aside for student presentations on the team projects. The teaching schedule on the module comprises three weekly lectures and one weekly tutorial. The lecture schedule is shared by two lecturers, whilst the tutorials are delivered by tutors, typically Graduate Teaching Assistants (GTAs).

[2] All students who study Quantitative Methods are also required to study Principles of Economics, which is the core principles course in Economics. We will use performance information from Principles of Economics to construct an alternative measure of ability.

The module starts in October of each year. Throughout the sample period the module was assessed by means of two tests, in Weeks 11 (January) and 21 (April) of term, one team project due in Week 25 (May), and a final exam in June. A timeline of the assessments is provided in Figure 1. The two in-term tests are each worth 12% of the final mark, the team project 16%, and the final exam 60% of the final mark. The two in-term tests are evaluated by the tutors, whilst the team project and the final exam are evaluated by the lecturers. Throughout the sample period, the overall course director is unchanged. Although the course content has marginally evolved over the time period under study, the assessment structure has remained the same.

[Insert Figure 1 about here]

Data are available at the individual student level for students taking the module in each of the 2007/08 to the 2012/13 academic years. Our data is a repeated cross-section over six years. We have data on each student's performance on each component of assessment. We also have information on students' gender, ethnic background, nationality, and entry qualification (data on entry grades are not available on a comparable basis). This allows us to construct our measures of diversity based on gender, nationality and ability. Table 1 presents the composition of the sample by year. Student numbers reflect the increasing intake of students into Economics and related subjects. In each year there are slightly more international students than female students: about one-third as compared with just over one-quarter. However, the percentage of international and female students remains fairly constant throughout the sample period.

2.2 Team Formation and Randomization

The basic premise of the experiment is that teams are exogenously treated to a random combination of individuals of different nationality, gender, and ability. We examine how

team composition impacts upon collective performance. Work teams in which all members of the group are from the same nationality, and of the same gender serve as our control group. The teams in the "treatment group" are subjected to different degrees of diversity, and our identification strategy relies on this variation. We also investigate how the treatment impacts upon their *ex post* individual performance.

To examine the true impact of team composition on performance, potential confounding factors need to be addressed. Individuals of similar abilities or skills could self-select into groups. Additionally, people with different demographic characteristics may tend to sort into different groups. Both of these would bias the estimate of the effect of team composition on collective performance. One way of circumventing these issues is for individuals to be exogenously and randomly assigned to the work-teams. In our setting, there are two stages of the randomization process.

In stage 1, at the beginning of the academic year, all enrolled students are allocated to tutorial groups of approximately 15 students by the course administrator. The process begins with a list of enrolled students in order of their enrolment. The allocation is random following an alphabetical algorithm of last names: the first student on the list is assigned to tutorial group 1, the second student to group 2, and so on. Each tutorial group is assigned a specific tutor throughout the course. In stage 2, project teams of 3 or 4 members are formed within each tutorial group. Each individual in a tutorial group is assigned to a project team by the course director on the basis of an alphabetical algorithm of last names. Similarly to the process in stage 1, the first student in the tutorial group is assigned to team 1, the second student to team 2, and so on. This two-stage process mitigates the risk of assortative and demographic sorting.[3]

[3] A potential concern with alphabetical algorithms is that people from certain countries may disproportionately have last names beginning with a handful of alphabets. For example many common Chinese surnames begin with X, Y and Z. This problem is mitigated to a large extent by the double randomization process.

The exogeneity of the group composition process may be compromised by movement of students between tutorial groups after the initial allocation for 'good reasons'. A good reason typically entails timetabling issues, sports commitments, caring responsibilities, *etc.*. Such moves typically occur for only 1 or 2 students per year. It is difficult to eliminate moves that are motivated by social networks between individuals. However they still have to be allocated to the same project team to influence the outcome. In sensitivity analysis we check the robustness of our results by excluding all project teams that include individuals who move tutorial groups. The results are very similar to those reported below.

If the randomisation has worked correctly, we should find no difference in *ex ante* characteristics between the treatment and control groups. Figure 2 shows the distribution of ability as measured by the sum of the marks from the two in-term tests. The distributions are very similar.

[Insert Figure 2 about here]

In Table 2, we compare the means of *ex ante* ability, team performance, and *ex post* individual performance (as measured by performance in the final exam) for teams with no diversity (the control group) and teams with varying degrees of diversity (the treatment group). 44 (27%) of the sample groups are composed of members of the same gender and same nationality, and they form the control group. There is no significant difference in means for *ex ante* ability and team performance, but individuals from the treatment group have higher *ex post* individual performance. Table 2 also shows that the team performance has a much smaller standard deviation than both measures of individual performance; this may be attributed to the different nature of the team task relative to the individual tasks.

[Insert Table 2 about here]

2.3 Incentives and Performance Outcomes

For the results of the field experiment to be economically meaningful, the participants should have strong individual and team incentives. A standard critique of using experiments in economics stems from weak incentives of the participants. In our setting, performance on the individual components of this module count towards the student's year-end grade, and students need to achieve a certain level of performance in order to progress to the second year of their undergraduate degree. The team project counts for 16% of their grade in Quantitative Methods, and all members of the team are assigned the same mark. Therefore, each member of the team has a natural incentive to contribute to the team project.[4]

For the outcomes of the team projects to be comparable, the project tasks need to be equivalent, if not the same. In our experimental setting, each project team is assigned a different task from a set of five standardized tasks. Since no more than five project teams can be formed within each tutorial group, no two project teams in the same tutorial group will be assigned the same task. The task is exogenously assigned to teams by the course director. The project entails performing some statistical analysis of a provided dataset, presenting the results, and submitting a joint project report based on the analysis. It assesses both quantitative and language/presentation skills. The intention is for teams to work independently on the task, without detailed supervision. The five tasks have remained the same throughout the study period, allowing comparison of different cohorts. A brief summary of the group tasks is provided in Appendix 2. All the tasks, although different in context, are similar in design and level of difficulty, and test the same skills: descriptive statistics, graphical representation, correlations, and bivariate regressions.

To ensure consistency in evaluation, all team projects are marked by the same individual -

[4] As with all team projects, it is possible for free-riding to occur. We do not have data on the interaction between members to control for this. Free-riding adds to the cost of co-ordination in a team, so may reduce group performance, but not *ex post* individual performance. In practice, since most groups have only 3 members, the extent of free-riding is relatively small, and on average affects only 1 or 2 groups in each year.

the course director. The project reports are anonymised, and graded on content, the quality of report-writing, and presentation skills. The course director has been the same individual over the sample period. Each project has objective and standardized marking criteria, and the final mark applies to all members of the group.

The design of our field experiment satisfies the conditions of exogenously formed teams, working on comparable tasks which are evaluated consistently. The main outcome of interest is whether team composition affects collective performance. *A priori*, one would expect teams to gain from diversity in the skill sets of individual members. To take the example of national diversity, emphasis on Mathematics and quantitative education at school varies by regions of the world. Since the task we use in our analysis has a large quantitative component, this may be a source of comparative advantage for some students. On the other hand, other students may have a comparative advantage in language skills, and hence in writing the report. Whilst it is possible that some students have an absolute advantage in many skills, the general idea is that a diverse team comprising individuals from different countries would, on average, have a wider skill set relative to teams without that diversity. A similar hypothesis can be formed about diversity in the gender of group members.

However, the gains from diverse skill sets may be offset by higher coordination costs of a diverse team. These coordination costs may be distinguished from the free-riding problem discussed in footnote 4. Apart from possible language barriers, there may also be social and cultural differences across group members. Some cultures are more egalitarian than others and the social norms of students from these countries may be quite different from those of students from more hierarchical cultures. All these put a constraint on the gains from diversity.

The main outcome of interest is the *ex-post* individual performance. Having been 'treated' to a diverse team, individuals may experience a gain in their subsequent performance. This may be due to diversified skills, or through enhanced motivation from having bench-marked

their ability to that of their team members. Also, subsequent individual performance is not subject to the coordination costs inherent in the team performance.

2.4 Economic Mechanisms

There are two parts to our analysis. In this subsection we provide an intuitive exposition of the mechanisms we have in mind; Appendix 6 presents a simple model which produces these results. First, we investigate the effect of team diversity on team performance. The task requires the use of a range of skills, including data manipulation and analysis, and both oral and written presentation. Students will differ in terms of their abilities in each of these skills. Therefore, a team with a more diverse skill-set, if it is able to utilise these skills effectively, would be expected to achieve a higher score in the group task. At the same time, a team with more diverse skills may also be more diverse in other dimensions, and this may make the team more difficult to manage; this is what Papps et al (2011) and Kahane et al (2013) find in professional sports. If the effort of managing the team takes away from the team's efforts towards completing the task, having a more diverse team may have a detrimental effect on team performance. Teams with wide dispersion of individual abilities do not tend to perform well. Therefore, whether or not a more diverse team results in better team performance depends on the degree of substitutability between team members (or how diverse they are), and the incremental cost of coordinating a more diverse team.

The second part of our analysis relates being in a diverse team to future individual performance. Here again the element of diversity is crucial. If your teammates are exactly the same as you, including having the same knowledge and skills, then there is nothing you can learn from your teammates. Therefore, it is only when your teammates are different from you, that there is something you can learn. As this suggests, learning is not instantaneous, but occurs over time. Learning can also be bidirectional. Often the assumption is that more-able

students can help less-able ones. This may be the case, but in the process of helping their colleagues, the more-able students may also enhance their own understanding, by having to organize their thoughts in order to communicate them.

These are, of course, simplifications. First, even if individuals have the same gender, nationality and measured ability, they may still be different from each other in some dimensions. We do not have data on other possible sources of heterogeneity, and we argue that controlling for the observed dimensions of diversity reduces the remaining diversity. Second, because the project teams are small (mostly three members to a team), we are unable to speak to the idea in the literature on professional sports (Papps et al, 2011, Kahane et al, 2013) that there may be an optimal degree of diversity. Third, we do not observe group interactions, so are unable to get a handle on the true mechanisms underlying the results observed. For example, we do not know if having participated in this group task, group members also discuss other tasks amongst themselves (for instance, other topics in the Quantitative Methods course). Similarly, we do not observe if teammates become friends as a result of this group task, and mutually support each other in the rest of their degree. These are possible channels through which the group task may influence future individual performance, but our data do not allow us to explore these possibilities. It is possible that any positive effect of group diversity on collective performance is exactly counterbalanced by the coordination costs. In that case, even without the ability to examine the mechanisms from observational data, we can detect some gains in longer term individual performance when coordination costs are absent.

3 METHOD AND RESULTS

3.1 Short-Term Treatment Effects

We estimate the short-term impact of diversity on team performance on the Quantitative Methods group project. That is, we estimate regressions of the following form, at the group level:

$$PG_{gt} = \rho A_{gt} + \boldsymbol{\beta}\mathbf{X}_{gt} + \alpha_t + \epsilon_{gt} \tag{1}$$

Where PG_{gt} is the performance of the team (the identical project mark awarded to all members of a given team), A_{gt} is prior student ability as measured by the sum of the marks attained in the two Quantitative Methods tests that precede the group project, $\mathbf{X}_{gt}$ is a vector of measures of diversity, α_t is a set of time dummies which control for unobserved effects that may be relevant (for instance, cohort effects), and ϵ_{gt} is a random error term.

We include measures of diversity for gender, nationality and ability. With teams of only three or four members each, we use relatively simple measures of diversity. First, we use a binary indicator, $AnyDiversity$, which equals 1 if the group is not composed of all British male students. In our sample, we have 44 groups (approximately 27%) with no diversity. Then, we construct a range of *de jure* measures of diversity.

The measure of gender diversity is the product of the percentage of male and female in a team:

$$D(G) = \%Male \times \%Female. \tag{2}$$

For a single-gender team, the measure of diversity will be zero, whereas the measure will be 0.25 when there are equal numbers of male and female members. The mean (median) $D(G)$

for our sample groups is 0.38 (0.29).

For nationality, we calculate the percentage of non-British students and the number of different nationalities represented in the group. The measure of diversity of nationality used is:

$$D(N) = \%Non\ British \times Number\ of\ Nationalities \tag{3}$$

If all team members are of the same nationality, $D(N) = 0$, and $D(N)$ takes a maximum value of 4 if all four group members are of different, non-British national origins[5]. The mean (median) $D(N)$ for our sample groups is 0.91 (1.05).

Our measure of diversity of ability is the standard deviation across group members of the ability measure (defined above as the sum of the marks from the two in-term tests)[6]:

$$D(A) = StdDev(Test1 + Test2) \tag{4}$$

Papps et al. (2011) show that the dispersion of individual ability in a team is a strong predictor of collective performance. In our specifications, we control for both average ability and the dispersion of ability. A team's average ability is the average of each member's ability score. The dispersion is measured as the difference between the highest and the lowest ability of group members.[7] The mean (median) $D(A)$ for our sample groups are 15.02 (18.04). Detailed summary statistics of the key variables are presented in Appendix 1.

Tables 3, 4, and 5 provide information on group composition by year, in terms of the per-

[5] As a robustness check, we also calculated a Herfindahl-type index as a measure of diversity of nationality. The two measures display a strong and positive correlation ($\rho = 0.62$).

[6] As a robustness check, we use the average of four preceding tests (two Quantitative Methods and two Principles of Economics) as a measure of students' *ex-ante* ability. The results are qualitatively similar for both measures of ability.

[7] We also check the robustness of of our results using the maximum ability rather than the average ability of the group.

centage of female, percentage of non-British students, and ability of students. As might be expected from Table 1, over half of all groups have no or only one female member, and similarly, over half of all groups have no or only one non-British member. However, in each year there is a significant percentage of groups which comprise primarily (and even occasionally exclusively) female or non-British members. The distribution of ability is somewhat normal, and there is no evidence that the proportion of students with high ability has increased significantly within the sample period.

[Insert Tables 3, 4, and 5 about here]

Table 6 presents the correlation matrix for the key variables used in the analysis. Individual performance in the final exam is strongly correlated with ability as measured by previous test scores. The three measures of diversity are only weakly correlated with each other. Figure 3 shows the effect of increasing team diversity on team and *ex post* individual performance. Increasing the number of nationalities has only a marginal effect on team performance, but a much larger monotonic effect on subsequent individual performance: as the number of nationalities in a team increases from 1 to 4, median individual performance increases by 23 percent, from 57 to 70.

[Insert Table 6 and Figure 3 about here]

The central hypothesis is that team diversity enhances *collective* performance. Using the measures of diversity discussed above, we estimate equation (1) to investigate whether a more diverse group leads to better group performance. The results are presented in columns (1) and (2) of Table 7. In column 1, we use a binary measure, $AnyDiversity$, of diverse groups. There is no statistically significant effect of this measure on group performance. In column 2, we use our constructed measures of diversity: $D(G)$ (gender diversity), $D(N)$ (nationality diversity) and $D(A)$ (diversity in ability). All these measures are not statistically

significantly associated with group performance at conventional levels. Therefore, there is no evidence to suggest that group diversity in terms of gender, nationality or ability, affects team performance. This may be explained by the productivity gains from diversity being offset by the increased coordination costs. However, the average ability of team members is positively related to team performance.

3.2 Long-Term Treatment Effects

Although we find no evidence that more diverse groups outperform less diverse ones, it may be that the gains from working in a diverse group can be manifest in subsequent individual performance. The proposition is that performance gains can result from weaker students learning from more able students. Alternatively, individuals may work harder after they become aware of other students' performance. Motivated by such considerations, we estimate equations of the following form:

$$PI_{it} = \psi A_{it} + \beta PG_{it} + \delta X_{it} + \alpha_t + \upsilon_{it} \tag{5}$$

Where PI_{it} is the *ex-post* individual performance, and the other variables are as previously described. The coefficient of interest is δ; controlling for prior individual ability and team performance, what is the impact of team diversity on future individual performance? A positive and significant estimate for δ would indicate that membership of a diverse team enhances subsequent individual performance. We also control for group performance and individual characteristics.

The results from estimating equation (5) are presented in columns (3) to (5) of Table 7. The dependent variable is individual performance in the Quantitative Methods final exam, which occurs after the team project. In column 3, we present the result with the binary measure $AnyDiversity$, which is positively and statistically significantly associated with subsequent

individual performance. In column 4, we present the results with our constructed measures of diversity. Gender and ability diversity in the group have no significant effects on final exam performance. However, nationality diversity in the group has a positive and significant effect on final exam performance. This suggests that even though the benefits of diversity are not manifest in the group performance, there is a spillover effect on subsequent individual performance from being a member of a group which is diverse in terms of nationality.[8] Although diversity has a positive impact on individual performance, the effect is small. In column (3), moving from a group with no diversity to a group with any diversity would increase individual performance by 0.03 percentage points, while in column (4) increasing $D(N)$ by 1 unit (about 1.4 standard deviations) would raise individual performance by 0.02 percentage points.

In columns 3 and 4, individual student performance in the final exam is positively associated with ability as measured by previous tests. It is also conceivable that students who has been a member of a higher-ability team does better in subsequent individual tasks. Therefore, in column 5, we control for the average ability of the group members for each student, in addition to their own ability. We find no statistically significant effect of the average ability of the group members over and above the positive effect of own ability. Overall, the results of Table 7 suggest that group diversity matters more for future individual performance than for contemporaneous group performance, and that the dimension of diversity that is most important is nationality. This suggests that the cultural differences associated with growing up in different countries is responsible for the individual learning from the experience in the group project.

[Insert Table 7 about here]

[8] A possible reason for no significant effect of gender diversity on performance is the low variation in the percentage of female members in groups. This is a typical result from the literature on gender diversity (see Adams and Ferreira, 2009 and Gregory-Smith et al. 2014), but is not the case in our sample; see Table 3.

A natural extension of the above enquiry is to examine the long-term persistence of individual performance gains. It may be argued that there is insufficient time between the team and subsequent individual tasks to allow us to comment on the persistence of any performance spillover from the group project. One way of overcoming this is to use the students' final degree outcomes as the measure of *ex post* individual performance. Degree outcomes are measured by the class of degree obtained, where a first class mark is 70 or above, an upper second class mark is 60 to 69, a lower second class mark is 50 to 59, a third class mark is 40 to 49, and any mark below 40 is a fail. Almost all students enrolled in Quantitative Methods complete their degrees in three years. We have data for three graduating cohorts: 2011, 2012 and 2013, corresponding to students who took Quantitative Methods in 2009, 2010 and 2011 respectively[9].

Making use of final degree outcomes as a measure of *ex post* individual performance introduces the danger of contaminating factors, since students may take different modules as part of their degree[10]. This choice of modules may in turn be influenced by assortative matching and selection by interest or ability. Therefore, we may be unable to identify a causal relationship between group diversity and degree outcomes. On the other hand, there is no reason to expect that any such confounding factors would have a different impact on members of diverse as opposed to homogeneous groups. We therefore interpret the following results as causal, but accept that the results may not be as clean as the previous results.

We use an ordered logit model to re-estimate equation (5) using final degree class as the dependent variable. This is a smaller sample because we do not have degree classifications of students who dropped out or moved to a different program. The results are presented in Table 8. In both specifications reported, having controlled for *ex ante* individual ability,

[9] In these three cohorts, 21 students (6.4% of the sample) either drop out, or switch to other degrees. We do not observe the long term outcome for these students.

[10] On the BSc in Economics, six of the eight second year modules are fixed (Intermediate Microeconomics and Macroeconomics 1 and 2, Mathematics, and Econometrics), but third (final) year modules can be chosen from a wide choice of Economics and related modules, with no compulsory modules.

individuals who were in first year teams that were more nationally diverse are more likely to have better degree outcomes at the end of their third year.[11] Note that, because we use an ordered logit in Table 8, the coefficients refer to the log odds, so a one unit increase in $D(N)$ results in a 0.05 increase in the log odds of attaining a higher degree class.

One remaining question is that of the mechanism that leads to better individual performance following an experience of working in diverse teams. From observational data it is not straightforward to observe the process of learning. The design of the course did not involve follow up surveys to examine these mechanisms. Therefore, we do not draw any inferences about the mechanisms from these results. To summarise, our results show that even though diverse teams do not outperform homogeneous teams in our field experiment, the benefits of diversity are manifest in enhanced subsequent performance of the team members. These better-performing individuals may enhance overall firm performance in real-world firms where group interactions are a repeated game.

[Insert Table 8 about here]

3.3 Additional Results

An additional question is whether the performance impact of diversity is different across different groups of individuals. For example, the gains from diversity may be limited to students with high (low) ability, or may differ between British and non-British students. To investigate these possibilities, we augment our specifications with the indicators *British* for British students, and *HighAbility* for students will high prior attainment. *HighAbility* equals one for students in the top decile of the ability score as defined in Section 3.1 above.[12]

[11] We checked the robustness of our baseline estimates using the smaller sample of students for which we have the degree classifications. Our results remain unchanged.

[12] We check the robustness of our results using different thresholds of *HighAbility* at 15%, 20% and 25% of the ability distribution.

We interact both $British$ and $HighAbility$ with the three measures of diversity. The results are presented in Table 9, where the dependent variables in columns (1) and (2) are team performance and subsequent individual performance respectively.

[Insert Table 9 about here]

In column (1), the coefficient on $British$ is negative and significant: controlling for average team ability, a team which comprises only British students has lower group performance, by 3.7 percent on average. However, all-British groups perform better when the groups are gender-diverse, as is evident from the positive and significant coefficient on $British * D(G)$. In column (2), we find evidence that British students perform less well than non-British students in the subsequent individual exam, by 5.5 percentage points, even though we have controlled for prior individual ability. Compared to the results above on the effect of diversity, this is a large coefficient. Since prior ability is measured by previous Quantitative Methods tests, this lower exam performance of British students cannot be explained by the idea that British students are weaker in Maths than non-British students. What it does suggest, is that British students perform worse in high-stakes situations, relative to students from other countries. There is also evidence that being exposed to a gender-diverse group enhances the subsequent individual performance of British students. None of the other interactions in Table 9 are statistically significant at conventional levels; our results do not differ by student ability.

In the main results, we constructed a measure of nationality diversity, $D(N)$, using the nationality of each team member. However, some of that diversity may have a sociocultural origin: some countries may exhibit similar personality traits. For example a group comprising a French, a Swiss and a Belgian student is less diverse than a group comprising a French, a Chinese and a Ghanaian student. We therefore check the robustness of our nationality measure by categorising nations into supra-national groups following Huntington's (1996)

classification. Since not all countries are represented in our sample, we use the following groups: Asian, African, Islamic, Orthodox and Western. A complete list of nationalities and their supra-national groups are presented in Appendix 3.

With this classification, we calculate a measure of supra-national diversity:

$$D(SN) = \%[Non - Western] \times Number\ of\ SupraNational\ Groups \tag{6}$$

So a group comprised of all European students has a value of 0 on this measure and a diverse group with no European students will have a higher score on the scale. Results using the supra-national groupings as a measure of diversity are provided in Table 10.

[Insert Table 10 about here]

In column (1) of Table 10, we find that groups which are supra-nationally diverse perform better than homogeneous groups. This is different from our baseline results, in which national diversity has no statistically significant impact on group performance. This suggests that what is important in terms of the impact of diversity on group performance is not merely that the group members are from different countries, but that they are from different sociocultural backgrounds. Similarly, we find in column (2) of Table 10 that the gain in subsequent individual performance from supra-national group diversity is larger than from national diversity in Table 7. Thus, individual workers gain more in subsequent individual productivity when they have been exposed to a work group comprising people from different sociocultural backgrounds. All other variables retain their previous sign and significance. In addition, the magnitude of the coefficients is similar to those in previous tables. Taken together, these results suggest that diverse work teams, particularly those composed of workers from different socioeconomic cultures, experience gains in both collective and individual productivity.

3.4 Robustness

We perform a series of robustness checks on our baseline results, the details of some of which are omitted in the interests of brevity. First, differences in team size may influence our results (see Biermann and Kearney, 2010). To ensure that this is not the case, we estimate our baseline models separately for 3- and 4-member groups. The results are very similar, although the estimates for 4-member groups are not statistically significant, probably because of the small sample of 4-member groups. The results are presented in Appendix 4.

Second, it may be argued that our control group is not homogeneous. An all-male team with members from the same nationality may differ in ethnicity, religion, and language spoken at home. Although there is information on ethnicity in our data, it is incomplete, as some students choose not to disclose this information. We do something slightly different. We define a different control group, which consists of teams where the members have the same gender and nationality (for instance, they may be a British all-female team, or a male all-Chinese team; in practice, there are only a few teams of this type). Therefore, our control group is a broader definition of a homogeneous group than in our baseline. The results are presented in Appendix 5, and are very similar to the baseline estimates. This suggests that our results are not driven by our definition of the control group in the main results as being a British all-male team.

We also conduct other robustness tests, the results of which are omitted for brevity. We estimate alternate specifications with dummies for no-foreigners, 1-foreigner, 2-foreigners to provide evidence on the extensive margin. We find that presence of any foreigner in a project team is associated with subsequent individual performance gains for all members of those diverse groups.

Next, whilst the course is compulsory for Economics majors, joint major students may self-select into it. This self-selection might affect performance through ability. Therefore, we

check the robustness of our results by using only the sample of Economics majors. This reduces the sample size, and the precision of the estimates, but the results are similar to the baseline estimates.

A standard critique of the experimental literature is the limited external validity of the results, because the artificial nature of the experiment means that it is difficult to draw general conclusions from the experimental results. Our setting, because it is not a laboratory experiment, has greater external validity, because the task performed is an actual, assessed task, and the participants are unaware of the experiment. Of course, because our setting occurs in a university environment, it is still difficult to generalise our results to the wider economy. Nonetheless, any organization can form task-teams with exogenous and random assignment of members, and in so doing replicate our experiment.

4 CONCLUSION

We use a field experiment to provide evidence on individual performance gains from working in a diverse team. Whereas team diversity has no significant effects on team performance, suggesting that the benefits of diversity may be counterbalanced by higher coordination costs of diverse team members, any skills assimilated from team-mates may be transferred to future individual tasks where the coordination cost no longer exists. The positive effect of team diversity is relatively small, and arises when teams are nationally diverse, but not when teams are diverse in terms of gender or ability. In addition, we find that a broader definition of sociocultural diversity has a positive and significant effect on both group and individual performance.

Our results contribute to the growing literature which suggests a possible mechanism through which diversity may affect group performance. Gains in the individual productivity of group

members resulting from working in diverse groups may translate into enhanced performance in future group tasks. Therefore, firms can benefit from higher overall labour productivity in the long run from employing diverse work teams.

REFERENCES

[1] Adams, R.B. and Ferreira, D., 2009. Women in the boardroom and their impact on governance and performance. *Journal of Financial Economics*, 94(2), pp.291-309.

[2] Bandiera, O., Barankay, I. and Rasul, I., 2010. Social incentives in the workplace. R*eview of Economic Studies*, 77(2), pp.417-458.

[3] Becker, G.S., 1974. A theory of marriage: Part II. *Journal of Political Economy*, 82(2, Part 2), pp.S11-S26.

[4] Carrell, S.E., Hoekstra, M. and West, J.E., 2015. *The impact of intergroup contact on racial attitudes and revealed preferences* (No. w20940). National Bureau of Economic Research.

[5] Chatman, J.A. and Flynn, F.J., 2001. The influence of demographic heterogeneity on the emergence and consequences of cooperative norms in work teams. *Academy of Management Journal*, 44(5), pp.956-974.

[6] Cox, T.H., Lobel, S.A. and McLeod, P.L., 1991. Effects of ethnic group cultural differences on cooperative and competitive behavior on a group task. *Academy of Management Journal*, 34(4), pp.827-847.

[7] Duflo, E., Glennerster, R. and Kremer, M., 2007. Using randomization in development economics research: A toolkit. *Handbook of Development Economics*, 4, pp.3895-3962.

[8] Ely, R.J., Padavic, I. and Thomas, D.A., 2012. Racial diversity, racial asymmetries, and team learning environment: Effects on performance. *Organization Studies*, 33(3), pp.341-362.

[9] Falk, A. and Ichino, A., 2006. Clean evidence on peer effects. *Journal of Labor Economics*, 24(1), pp.39-57.

[10] Green, C.P. and Homroy, S., 2018. Female directors, board committees and firm performance. *European Economic Review*, 102, pp.19-38.

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

[12] Huntington, Samuel P., 1996. The clash of civilizations and the remaking of world order, *Simon & Schuster,* New York.

[13] Kahane, L., Longley, N. and Simmons, R., 2013. The effects of coworker heterogeneity on firm-level output: assessing the impacts of cultural and language diversity in the National Hockey League. *Review of Economics and Statistics*, 95(1), pp.302-314.

[14] Katz, L.F., Kling, J.R. and Liebman, J.B., 2001. Moving to opportunity in Boston: Early results of a randomized mobility experiment. *Quarterly Journal of Economics*, 116(2), pp.607-654.

[15] Kochan, T., Bezrukova, K., Ely, R., Jackson, S., Joshi, A., Jehn, K., Leonard, J., Levine, D. and Thomas, D., 2003. The effects of diversity on business performance: Report of the diversity research network. Human Resource Management, 42(1), pp.3-21.

[16] Lazear, E.P., 1999. Globalisation and the market for team-mates. *Economic Journal*, 109(454), pp.15-40.

[17] Lee, N., 2015. Migrant and ethnic diversity, cities and innovation: Firm effects or city effects?. *Journal of Economic Geography*, 15(4), pp.769-796.

[18] Nathan, M., 2013. Top Team Demographics, Innovation and Business Performance: Findings From English Firms And Cities, 2008-9.

[19] Nathan, M. and Lee, N., 2013. Cultural Diversity, Innovation, and Entrepreneurship: Firm-level Evidence from London. *Economic Geography*, 89(4), pp.367-394.

[20] Nathan, M., 2015. Same difference? Minority ethnic inventors, diversity and innovation in the UK. *Journal of Economic Geography*, 15(1), pp.129-168.

[21] O'Reilly III, C.A., Caldwell, D.F. and Barnett, W.P., 1989. Work group demography, social integration, and turnover. *Administrative Science Quarterly*, pp.21-37.

[22] Papps, K.L., Bryson, A. and Gomez, R., 2011. Heterogeneous worker ability and team-based production: Evidence from major league baseball, 1920–2009. *Labour Economics*, 18(3), pp.310-319.

[23] Polzer, J.T., Milton, L.P. and Swarm Jr, W.B., 2002. Capitalizing on diversity: Interpersonal congruence in small work groups. *Administrative Science Quarterly*, 47(2), pp.296-324.

[24] Rockenbach, B., Sadrieh, A. and Mathauschek, B., 2007. Teams take the better risks. *Journal of Economic Behavior & Organization*, 63(3), pp.412-422.

[25] Sacerdote, B., 2001. Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics*, 116(2), pp.681-704.

[26] Smith, A., 1776. An inquiry into the nature and causes of the wealth of Nations, *W. Strahan and T. Cadell*, London.

[27] Tomaskovic-Devey, D., Zimmer, C., Stainback, K., Robinson, C., Taylor, T. and McTague, T., 2006. Documenting desegregation: Segregation in American workplaces by race, ethnicity, and sex, 1966–2003. *American Sociological Review*, 71(4), pp.565-588.

[28] Vogel, R., Puhan, T.X., Shehu, E., Kliger, D. and Beese, H., 2014. Funding decisions and entrepreneurial team diversity: A field study. *Journal of Economic Behavior & Organization*, 107, pp.595-613.

[29] Watson, W.E., Kumar, K. and Michaelsen, L.K., 1993. Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups. *Academy of Management Journal*, 36(3), pp.590-602.

[30] Zarraga, C. and Bonache, J., 2005. The impact of team atmosphere on knowledge outcomes in self-managed teams. *Organization Studies*, 26(5), pp.661-681.

[31] Zenger, T.R. and Lawrence, B.S., 1989. Organizational demography: The differential effects of age and tenure distributions on technical communication. *Academy of Management Journal*, 32(2), pp.353-376.

Figure 1

Timeline of the Key Stages of the Field Experiment

This figure places the key stages of the field experiment on a timeline. The combined performance in the January and April tests serves as a measure of *ex ante* ability before the group project in April and May. The performance of individual group members in the final exam in June allows us to examine the impact of group diversity on *ex post* individual performance.

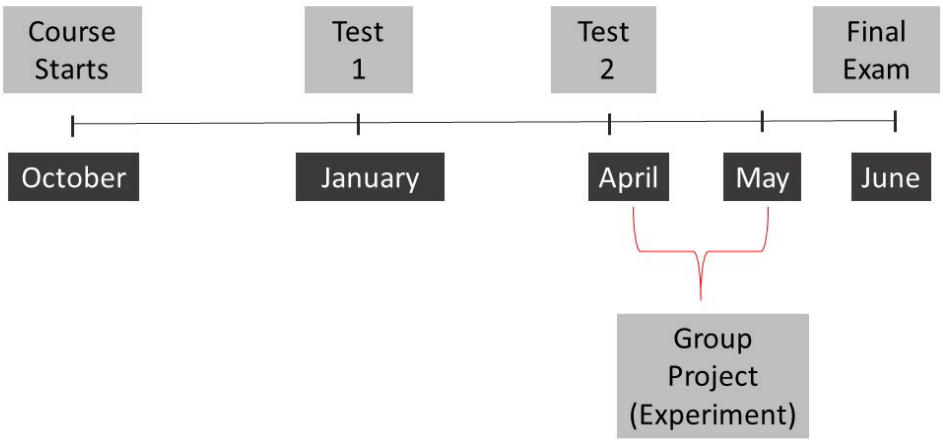

Figure 2

Distribition of Ability for Treatment and Control Groups

This figure shows the distribution of the *ex ante* ability of individuals assigned to the treatment group (teams with some levels of diversity), and control group (teams with members from the same nationality and gender). The distributions are similar, with the treatment group having a slightly higher mean than the control group. We test for the difference of means in Table 2.

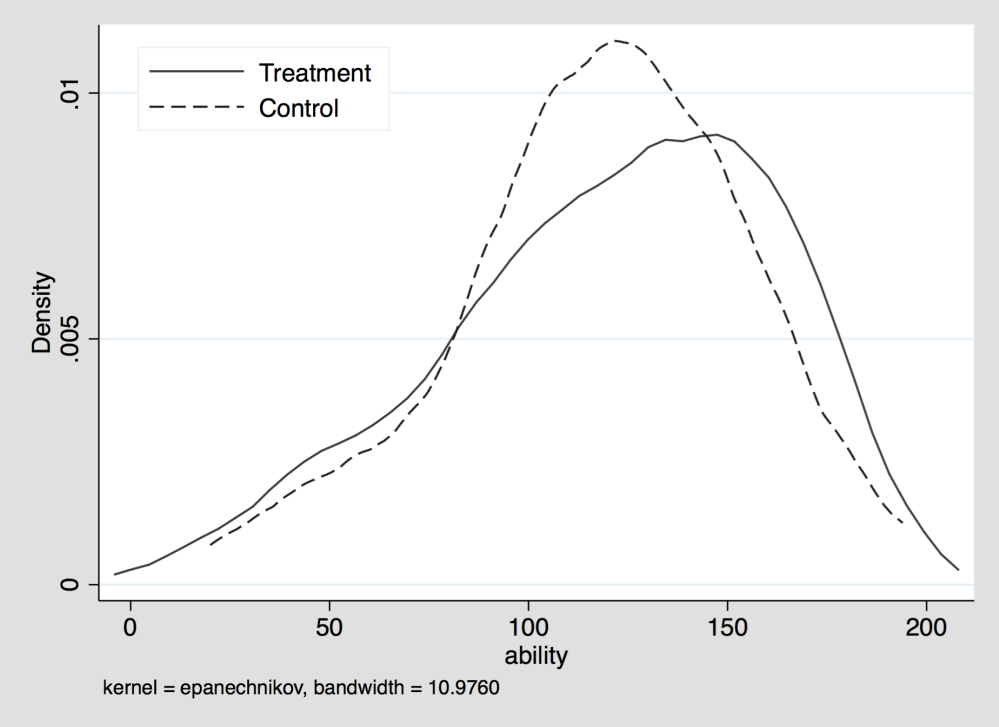

Figure 3

Variation in Team Diversity and Performance

This figure shows the effect of different levels of national diversity on mean group performance and *ex post* individual performance. Group performance is more similar to a normal distribution whereas individual performance is skewed.

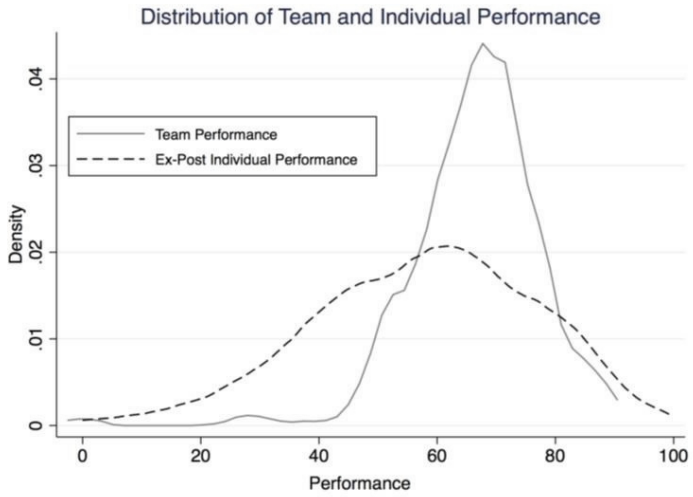

Table 1

Descriptive Statistics on Sample Diversity

This table presents the diversity of the sample. The number of students and groups have steadily increased over time. No particular trends are observed in the proportion of international or female students within our sample period.

Year	No. of groups	No. of students	% International Students	% Female Students
2008	15	51	37.25	29.41
2009	17	54	31.48	24.07
2010	28	96	28.12	27.08
2011	34	105	30.48	22.86
2012	32	107	42.45	30.18
2013	37	117	34.18	31.62

Table 2

Comparison of Means

In this table we present univariate comparisons of teams with no diversity (control), and teams with some levels of diversity (treatment). Control groups have all members from the same gender and nationality. *Ex ante* ability is measured on a scale from 0 to 200, while group performance and *ex post* individual performance are measured on a scale from 0 to 100. There is no significant difference between treatment and control groups in *ex ante* ability and team performance. p-values are given in brackets.

	Treatment	Control	Difference
N	68	44	-
Ex-Ante Ability	130.35	126.99	3.36 (0.194)
Team Performance	66.31	66.80	-0.49 (0.239)
Ex-Post Individual Performance	58.04	51.98	6.06** (0.001)

Table 3

Gender Diversity and Group Composition

This table presents the variation in group gender diversity over the sample period. A group may consist of 3 members, in which case the percentage of females can take values equal to 0, 33, 67 or 100%; or 4 members, in which case the percentage of females can take values of 0, 25, 50, 75 or 100%.

Percentage of Female in groups	2008	2009	2010	2011	2012	2013
	% Female					
0	35.29	29.63	34.38	42.86	20.56	43.59
25	15.69	42.59	16.67	7.62	26.17	6.84
33	17.65	11.11	18.75	40	22.43	14.53
50	19.61	11.11	20.83	3.81	7.48	6.84
67	5.88	5.56	9.38	5.71	19.63	20.51
75	0	0	0	0	3.74	3.42
100	5.88	0	0	0	0	4.27

Table 4

National Diversity and Group Composition

This table presents the variation of diversity of nationality in groups over the sample period. A group may consist of 3 members, in which case the percentage of non-British students can take values equal to 0, 33, 67 or 100%; or 4 members, in which case the percentage of non-British members can take values of 0, 25, 50, 75 or 100%.

Percentage of Non-UK students in groups	2008	2009	2010	2011	2012	2013
	% Students					
0	25.49	37.04	30.21	40	22.43	35.9
25	15.69	20.37	20.83	3.81	14.95	9.4
33	23.53	11.11	21.88	25.71	14.02	23.08
50	15.69	0	20.83	9.52	14.95	6.84
67	5.88	16.67	3.12	11.43	19.63	11.97
75	7.84	14.81	0	3.81	11.21	0
100	5.88	0	3.12	5.71	2.8	12.82

Table 5

Ability Diversity and Group Composition

This table presents the variation of ability diversity in groups over the sample period. The measure of Ability is the sum of Test1 and Test2 grades. We provide the fraction of students for the ranges: 0 - 50, 51 - 100, 101 - 150 and 151 - 200.

Ability	2008	2009	2010	2011	2012	2013
			% Students			
0 - 50	18.21	17.42	19.05	20.09	20.17	22.31
51 - 100	29.33	30.76	34.67	35.22	34.98	35.09
101 - 150	21.58	34.58	28.93	26.17	34.54	29.67
151 - 200	15.19	17.24	17.35	18.52	10.31	12.93

Table 6

Correlation of Key Indicators

In this table we present the correlation between the key diversity indicators $D(G)$ (gender diversity), $D(N)$ (national diversity) and $D(A)$ (diversity in ability), and the performance indicators. Gender diversity $D(G)$ is positively associated with nationality diversity $D(N)$. Ability is positively associated with both group and individual performance. $D(N)$ is also positively associated with individual performance. Associations which are statistically significant at the 1% level are marked by *.

	$D(G)$	$D(N)$	$D(A)$	Ability	Group Perf.	Individual Perf.
$D(G)$	1.000					
$D(N)$	0.2752*	1.000				
$D(A)$	0.0722	-0.0623	1.000			
Ability	-0.0184	0.0932*	-0.1612*	1.000		
Group Performance	0.0253	-0.0053	0.0633	0.1171*	1.000	
Individual Performance	-0.0277	0.1367*	-0.0814	0.7635*	0.1055	1.000

Table 7

Effect of Diversity on Group and Individual Performance

In this table, we present the effect of diversity in groups on contemporaneous group performance and *ex-post* individual performance. $AnyDiversity$ is an indicator for groups that are not composed solely of male British students. $D(A)$ refers to diversity in ability, $D(G)$ to gender diversity, and $D(N)$ to diversity in nationality. In columns (1) and (2) each observation is a group, while in columns (3) to (5) each observation is an individual. Robust standard errors in parentheses. *,** ,and *** refer to statistical significance at 10%, 5% and 1% respectively.

Dependent Variable	Group Performance		Ex-post Individual Performance		
	(1)	(2)	(3)	(4)	(5)
Ability	0.021**	0.034**	0.304**	0.353***	0.228**
	(0.010)	(0.012)	(0.024)	(0.013)	(0.091)
Average Ability-					0.007
Group Members					(0.003)
AnyDiversity	0.077		0.028**		
	(0.042)		(0.012)		
$D(A)$		0.040		0.000	0.000
		(0.047)		(0.038)	(0.001)
$D(G)$		0.000		-0.000	-0.000
		(0.000)		(0.000)	(0.000)
$D(N)$		-0.001		0.0179**	0.0182**
		(0.008)		(0.006)	(0.011)
$GroupPerformance$			0.028	0.065	0.057
			(0.017)	(0.063)	(0.051)
Adjusted R^2	0.191	0.203	0.415	0.620	0.608
Year Dummies	Yes	Yes	Yes	Yes	Yes
Observations/Clusters	163	163	530	530	530

Table 8

Long Term Performance Impact

In this table we present the longer term impact of team diversity on individual performance. In both columns, degree classification is the dependent variable. The reported coefficients are from ordered logit estimation. $D(A)$ refers to diversity in ability, $D(G)$ refers to gender diversity, and $D(N)$ refers to diversity in nationality. Robust standard errors in parentheses. *,** , and *** refer to significance at 10%, 5% and 1% levels respectively.

Dependent Variable	Degree Classification	Degree Classification
	(1)	(2)
$D(A)$	0.031	0.027
	(0.023)	(0.019)
$D(G)$	0.024	0.022
	(0.017)	(0.015)
$D(N)$	0.056*	0.049*
	(0.030)	(0.25)
$British * D(A)$		0.003
		(0.007)
$British * D(G)$		0.000
		(0.001)
$British * D(N)$		0.009
		(0.016)
Pseudo- R^2	0.189	0.466
Year Dummies	Yes	Yes
Observations	308	308

Table 9

Differential Gains from Team Diversity

We investigate if particular groups of individuals, viz. British students and high-ability students, gain more from diversity. There is no strong evidence of differential gains from diversity for any particular group. $D(A)$ refers to diversity in ability, $D(G)$ refers to gender diversity, and $D(N)$ refers to diversity in nationality. In column (1) each observation is a group and in column (2) each observation is an individual. Robust standard errors in parentheses. *,** ,and *** refer to significance at 10%, 5% and 1% levels respectively.

Dependent Variable	Group Performance	Ex-post Individual Performance
	(1)	(2)
Ability	0.023*	0.357***
	(0.012)	(0.023)
$D(A)$	0.0127	0.025
	(0.042)	(0.051)
$D(G)$	-0.001	-0.001*
	(0.006)	(0.000)
$D(N)$	0.028	0.016**
	(0.009)	(0.007)
British	-3.694	-5.539*
	(2.475)	(3.018)
$British * D(A)$	0.0417	0.0514
	(0.050)	(0.060)
$British * D(G)$	0.002**	0.0019*
	(0.000)	(0.001)
$British * D(N)$	0.0019	0.0053
	(0.012)	(0.0153)
$HighAbility * D(A)$	-0.147	0.126
	(0.100)	(0.090)
$HighAbility * D(G)$	-0.000	0.000
	(0.001)	(0.001)
$HighAbility * D(N)$	0.011	0.019
	(0.016)	(0.019)
Group Performance		0.052
		(0.059)
Adjusted R^2	0.232	0.640
Year Dummies	Yes	Yes
Observations/Clusters	163	530

Table 10

Supra-National Diversity and Performance

Since some countries are more similar than others, we present the results of the effect of supra-national diversity on group and ex-post individual performance. Supra-National diversity enhances both group and individual performance. $D(A)$ refers to diversity in ability, $D(G)$ refers to gender diversity, and $D(SN)$ refers to supra-national diversity. In column (1) each observation is a group and in column (2) each observation is an individual. Robust standard errors in parentheses. *,**,and *** refer to significance at 10%, 5% and 1% levels respectively.

Dependent Variable	Group Performance	Ex-post Individual Performance
	(1)	(2)
Ability	0.0234	0.3491***
	(0.0144)	(0.0136)
$D(A)$	-0.0016	-0.0466
	(0.0652)	(0.0434)
$D(G)$	-0.0004	-0.0013
	(0.0010)	(0.0008)
$D(SN)$	0.0262*	0.0192*
	(0.0136)	(0.0108)
$British$	-4.4155	-6.4464**
	(2.7489)	(2.8481)
$British * D(A)$	0.0519	0.0612
	(0.0654)	(0.0602)
$British * D(G)$	0.0017*	0.0018*
	(0.001)	(0.001)
$British * D(SN)$	-0.0029	0.0085
	(0.0191)	(0.0079)
$HighAbility * D(A)$	-0.1827	0.1045
	(0.113)	(0.0866)
$HighAbility * D(G)$	0.000	0.0029
	(0.0016)	(0.0017)
$HighAbility * D(SN)$	-0.0054	0.0233
	(0.0361)	(0.0406)
Group Performance		0.0433
		(0.0347)
Adjusted R^2	0.149	0.582
Year Dummies	Yes	Yes
Observations/Clusters	163	530

Appendix 1

Summary Statistics of Key Variables

In this table, we present the summary statistics of the key variables used in our regression analyses. $D(A)$ refers to diversity in ability; $D(G)$ refers to gender diversity and $D(SN)$ refers to supra-national diversity.

	N	Mean	Median	Std. Dev.
$D(A)$	163	15.02	18.64	10.07
$D(G)$	163	0.38	0.29	0.16
$D(N)$	163	0.91	1.05	0.72
$D(SN)$	163	0.63	0.75	0.48
Group Performance	163	64.17	67.92	17.26
Ability	530	58.44	53.56	22.64
Individual Performance	530	66.77	45.45	28.05

Appendix 2

Description of Group Projects

A brief description of the team projects is presented here. The data for all projects were provided by the course director. All tasks are standardized to similar levels of difficulty. All projects involve presentation of descriptive statistics, correlation matrices, and bivariate regressions. The students are expected to comment on the results.

Project Title	Frequency (All years)	Brief Description of the Project
Currency Crisis, Trade, and Contagion	33	Using supplied data, students examine the association between crises, trade volume, and contagion effects.
UK House Prices	36	Time-series analysis of UK house prices, and identification of bubbles using graphical methods.
Economic Growth	29	Panel data of analysis of GDP growth, and indications of convergence.
Zipf's Law and Urban Growth	31	Testing Zipf's law and urban growth on a sample of Malaysian cities.
Inflation in the United Kingdom	34	Time series analysis of inflation in the UK, with comments on recessions.

Appendix 3

Nationality and Supra-National Groupings

The sample contains students from only the listed countries. Using the definitions in Huntington (1996), we classify the nationalities into the following sociocultural groups: African, Asian, Islamic, Orthodox, and Western. Some sociocultural groups from the original Huntington (1996) classification may not be represented here.

Country	Students (All years)	Supra-National Grouping	Country	Students (All years)	Supra-National Grouping
Tanzania	01	African	Ukraine	01	Orthodox
Ivory Coast	01	African	Romania	01	Orthodox
Nigeria	04	African	Russia	02	Orthodox
China	46	Asian	Poland	07	Western
Malaysia	05	Asian	Ireland	03	Western
Mongolia	02	Asian	Spain	05	Western
Hong Kong	07	Asian	Greece	08	Western
India	05	Asian	Czech Republic	02	Western
Nepal	01	Asian	Cyprus	07	Western
Vietnam	03	Asian	France	05	Western
Pakistan	04	Islamic	Italy	02	Western
Kazakhstan	05	Islamic	Finland	01	Western
Indonesia	01	Islamic	Norway	04	Western
UAE	01	Islamic	Denmark	01	Western
Bahrain	02	Islamic	Germany	08	Western
Somalia	01	Islamic	Switzerland	01	Western
Turkmenistan	01	Islamic	Latvia	01	Western
Lithuania	16	Orthodox	Sweden	03	Western
Estonia	02	Orthodox	United Kingdom	211	Western
Bulgaria	08	Orthodox			

Appendix 4

Effect of Team Diversity in three member Teams

We limit the sample to teams with three members only. The results are similar to those in Table 6 for the full sample. $D(A)$ refers to diversity in ability, $D(G)$ refers to gender diversity, and $D(N)$ refers to diversity in nationality. In column (1) each observation is a group and in column (2) each observation is an individual. Robust standard errors in parentheses. *,** ,and *** refer to significance at 10%, 5% and 1% levels respectively.

Dependent Variable	Group Performance	Ex-post Individual Performance
	(1)	(2)
Ability	0.022**	0.356***
	(0.010)	(0.018)
$D(A)$	0.038	0.000
	(0.026)	(0.033)
$D(G)$	0.000	-0.000
	(0.000)	(0.000)
$D(N)$	-0.001	0.015**
	(0.004)	(0.004)
Group Performance		0.0522
		(0.049)
Adjusted R^2	0.194	0.581
Year Dummies	Yes	Yes
Observations/Clusters	151	482

Appendix 5

Effect of Diversity: Alternate definition of the control group

In this table, we present the effect of diversity in groups on group and individual performance with respect to control groups composed of students of the same gender and nationality (e.g. they may be all-female, or from a country other than Britain). Robust standard errors in parentheses. *,** , and *** refer to significance at 10%, 5% and 1% levels respectively.

Dependent Variable	Group Performance	Ex-post Individual Performance
	(1)	(2)
Ability	0.025**	0.304**
	(0.011)	(0.024)
Any Diversity	0.091*	0.038**
	(0.051)	(0.014)
Group Performance		0.019
		(0.013)
Adjusted R^2	0.210	0.428
Year Dummies	Yes	Yes
Observations/Clusters	163	530

Appendix 6: Theoretical model

Consider the following simplified environment, in which there are two students, $i = a, b$, who can each be of one of two types, 1 and 2 (the results generalise to many students of many types). Each student devotes H_a and H_b hours to the group project, and their productivity is A_a and A_b, respectively. The production function of the group task is:

$$Q = (\sum A_{1i}H_{1i})^{\sigma} + (\sum A_{2i}H_{2i})^{\sigma} \tag{7}$$

where $0 < \sigma < 1$ is the degree of substitutability between *types* of students. As usual, the larger is σ, the greater the degree of substitutability between types of students. If the two students are of the same type, there are no coordination costs, and therefore output of the non-diverse group is:

$$Q_{ND} = (A_a H_a + A_b H_b)^{\sigma} \tag{8}$$

If the two students are not of the same type, there is a coordination cost, which depends on the substitutability between students. Suppose that this cost reduces the hours devoted to the group task by the fraction $\sigma^{\psi} < 1$. Then, output of the diverse group is:

$$Q_D = (\sigma^{\psi} A_a H_a)^{\sigma} + (\sigma^{\psi} A_b H_b)^{\sigma} \tag{9}$$

Suppose that the two students have the same ability, and devote the same number of hours to the group task. That is, $A_a = A_b$ and $H_a = H_b$. First, note that, in the absence of a coordination cost, $Q_D > Q_{ND}$ as long as $\sigma < 1$ (i.e. different types of students are not perfect substitutes). Then, in the presence of a coordination cost, $Q_D > Q_{ND}$ if:

$$2 > \sigma^{\frac{\psi\sigma}{\sigma-1}} \tag{10}$$

That is, the diverse group will have higher output (i.e. better performance) than the non-diverse group if students are less substitutable (σ is small), and/or the coordination cost ψ is sufficiently small.

It is then a small step from this model, to consider the implications for individual learning from the group task. If the two students are of the same type, no learning occurs, so $A_{i,t+1} = A_{i,t}$. On the other hand, if the two students are of different types, then they can each learn from the other, and the amount of learning may depend on how different the two types of students are from each other, so $A_{i,t+1} = A_{i,t} + \frac{\alpha}{\sigma^{\delta}}$, where $\delta < 1$. Here, the smaller is σ, the lower the substitutability across students, hence the more learning that can take place.