

Group size and the use of skills diversity in production.

Brais Álvarez Pereira*

Shan Aman-Rana[†]

November 15, 2018

Brais Álvarez Pereira's Job Market Paper

Abstract

In this paper we analyse the relationship between group size and the effect of skills diversity on workers' performance. We explore this link theoretically in a model with knowledge spillovers in production, which predicts that the effect of diversity on performance increases with group size. We test this hypothesis in a lab experiment in Guinea-Bissau. Nursing and Economics students solve an individual test twice, first in isolation and then communicating with other participants, in homogeneous and diverse groups of size two and four. The average gain in the number of correct answers between both tests is the same for a Nursing student in a homogeneous or a diverse small group. For Nursing students placed with an Economics student in a large group, this gain is twice the average gain for Nursing students in large homogeneous groups. We also provide evidence in support of this mechanism from a novel administrative dataset of tax collectors in the Pakistan Administrative Services. Using a policy requirement whereby 10% of the civil servants must come directly from the army to observe diversity, we find that having an ex-military colleague is associated with a larger performance for civilian tax collectors the larger the group in which the ex-military member is placed. Our findings indicate that group size plays a central role in determining both the dimension and the sign of the effect of skills diversity on performance.

JEL Classification: C92, D00, D23, D73, D89, D91, O12

Keywords: skills diversity, group composition, productivity, experimental economics, organizational economics, public sector, development.

*Guinea-Bissau Government, Ministry of Economy and Finance, Overseas Development Institute Fellow, brais.alvarespereira@gmail.com

[†]London School of Economics, Economics department.

1 Introduction

*"... the problem of what is the best way of utilizing knowledge initially dispersed among all the people is at least one of the main problems of economic policy —or of designing an efficient economic system". F.A. Hayek 1945, *The Use of Knowledge in Society*.*

One of the main purposes of economic organizations is to coordinate the solution of the variety of tasks associated to production in the presence of workers' diverse skills and knowledge (Garicano, 2000), given individuals' capacity to specialize in solving a limited number of problems. When the production task is complex or knowledge-intensive, such that an objectively correct solution exists but a single worker does not always know it, matching individual workers and problems is not trivial. Efficiency in production requires workers' allocation to maximize knowledge complementarity subject to communication costs.

In this paper we propose and test a mechanism that explains how workers with diverse skills can be efficiently combined into groups, such that productivity is maximized. We provide causal evidence from a lab experiment run in Guinea-Bissau, and draw stylized facts from the field using administrative data on the performance of Pakistani tax collectors.

We use the main elements from the theoretical model on knowledge spillovers and group production in Lazear (1999b) to explore the effect of diversity on performance in groups of different size. Each worker solves an individual complex task and can improve her individual performance by talking to colleagues who are potentially better than her at solving it.¹ The knowledge of similar workers in the group becomes more and more redundant as their number increases, implying that the marginal contribution of a same-type worker to her colleagues' expected performance is decreasing in group size. This implies that the relative contribution of a fixed number of diverse² workers to their colleagues' performance is increasing in group size. This is a very important result because it means that group size does not only determine the size of the effect of diversity on performance, but it might also change its sign.

We test this prediction by running a lab experiment in Guinea-Bissau. We instruct Nursing and Economics students —nurses and economists from now on —to solve a test composed of 25 multiple-choice questions in 30 minutes, first in isolation and then communicating in groups of size 2 and 4. Nurses are the treated participants, with group size and economists providing a source of exogenous variation in available knowledge. We measure performance as the improvement in the number of correct answers between the individual and the group test. For nurses in groups of different size, we compare the improvement in test performance in homogeneous groups that only have students of the same degree relative to those that were randomly assigned to a group with one economist. Two things to note: first, average performance is the same for nurses in small and large homogeneous groups. Second, the null contribution of the 4th nurse to her colleagues' performance implies that it will be much better to place an economist in this position than in a small group. In groups of size 2 being paired with a nurse is as good for performance as being

¹Adapting the existing theoretical models on the relationship between skills diversity, knowledge spillovers and group performance to the individual level allows us to derive predictions which, with a careful experimental design, we can test without the dangers of ecological fallacies and other identification problems associated with the empirical inference of group effects (Manski, 1993; Angrist, 2014).

²We are focusing on the case in which diversity matters through differences in skills or knowledge. Heterogeneous within-type and between-type costs of communication in the theoretical model make our findings also relevant for the study of discrimination (differences in ingroup and outgroup behavior), typically associated to ethnic, religious, gender or national diversity (Alesina and La Ferrara 2000, 2005; Bertrand and Duflo 2017).

paired with an economist. In groups of size 4, the performance of nurses who can communicate with an economist and two other nurses is twice (97.3% higher) that of nurses in homogeneous groups. An extra treatment with degree and gender diverse groups shows causal evidence for apparently non-rational discrimination towards women, suggesting an interesting avenue for future research.

The wide applicability of our findings is corroborated with evidence from a dataset of Pakistani tax collectors. Each observation corresponds to an Assistant Commissioner (AC) in charge of the administration of one tehsil (municipality), in a given month. The measure of performance is given by the amount of taxes collected in that month by the team of lower-rank revenue officials she coordinates, over the annual target. Tehsils are grouped into districts, with those high-ranked civil servants working in the same district having periodic frequent meetings that create the potential for knowledge spillovers. Those elite civil servants in the same district at the same time will define a group. A rule introduced by the military in 1980 implies that a small proportion of the civil servants working in these districts come from the army. The data shows that, as predicted by our model, the performance of civilian ACs with an ex-military colleague in their group is higher the larger the size of the group.³

These results suggest that group size plays a central role in the relationship between skills or knowledge diversity and workers' performance. Our findings contribute to the literature in several ways. First, we are the first to provide causal and individual-level evidence on a direct implication of a well-established theory: the powerful theoretical models on skills diversity and production that have inspired our work, specifically designed to explain phenomena such as the structure of global firms, trading, migration or linguistic patterns (Lazear, 1998, 1999b,a; Antràs et al., 2006a,b) had been indirectly tested, using aggregate data at the country or sub-national level. Second, previous empirical and experimental papers (Haltiwanger et al., 1999; Hansen et al., 2006; Hoxby, 2000; Gagliarducci and Paserman, 2011) have explored the effect of different kinds of diversity (gender, age, race, etc.) on performance,⁴ but had neglected the role of group size. We close this gap.

The capacity of group size to change the sign of the effect of concrete kinds of diversity on workers' performance implies that our findings have the potential to help explaining existing conflicting observations.

The paper is organized as follows: Section 2 develops the theoretical model both for homogeneous and diverse groups. Section 3 explains the main elements of the lab experiment in Guinea-Bissau and its results. Section 4 describes in detail the field situation in Pakistan, the data, the empirical models used to test the corresponding hypotheses, and reports the main findings. Section 5 concludes.

2 The theoretical model

We use the main elements from the theoretical model in Lazear (1999b) to explore the role that group size plays in the relationship between diversity and performance. A complex task will be of one or another class depending on the state of the world. Two types of workers with different knowledge or skills can each solve one of the tasks, but not the other.

³In the appendix we extend the model to account for these dynamics in more complex settings, with interesting predictions found to hold in a hierarchical interpretation of these groups.

⁴See also Williams and O'Reilly III 1998; Pelled et al. 1999 for two extensive reviews of the Management and Sociology literature on the topic.

Conditionally on one type of workers being able to solve the task at hand, there is a distribution of talent across them: some are more skillful than others. Through communicating with other group-members, agents become as good in solving the corresponding task as the best colleague they have talked to. Communication is costly and can be interpreted as time far from production. Talking to a diverse colleague is more difficult or costly than talking to a similar one. This can be interpreted as a consequence of differences in jargon and technical knowledge, not easy to communicate to people from a different background.⁵

The main change we introduce to the original model, using it to explain individual instead of group production is trivial: it only requires assuming that each individual becomes as good as the best colleague she talks to, instead of having group production being implemented by the best worker. Hence, we do not really introduce any innovative element into the model.⁶ What we do in this section is to explore the role of group size in detail. Lazear’s paper explores the optimality of having diverse as compared to homogeneous groups of size 2. Our focus here is to explore how the effect of diversity changes as the number of similar workers in the group increases.

When we explore the role of diversity in groups with a different number of similar workers, the model produces a clear prediction: the effect on colleagues’ performance of introducing a fixed number of diverse worker into an otherwise homogeneous group is increasing in group size. This prediction will allow us to empirically test the concavity of knowledge spillovers in group size, the main element driving the results in the model.

2.1 Set up

Suppose a firm is hiring workers to transcribe the responses to a survey about the use of medical practices and medicines in a country’s health-centers, from hand-written paper questionnaires to electronic format. Each worker transcribes a questionnaire Z , a random variable that requires health-specific knowledge with probability p (task Z_1) and specialized knowledge of the statistical software with probability $1 - p$ (task Z_2).

The pool of candidates for the job is composed of ex-ante identical Nursing students and ex-ante identical Economics students —nurses and economists. Nurses (type-1 workers) have a stronger background in the relevant health-related issues and economists (type-2) in the use of the statistical software. To get intuition about the role of diversity in the model, we consider the extreme case in which the knowledge of nurses and economists is fully disjoint: a nurse never knows the solution to a software task Z_2 and an economist never knows the solution to a health task Z_1 . The background needed for the transcription of each questionnaire is not ex-ante known by the manager, implying that both nurses and economists have the same probability of facing task Z_1 and task Z_2 .

A worker’s performance when transcribing one questionnaire in isolation will be simply determined by the task at hand and her type and ability. Conditional on the task being of class Z_1 , type-1 worker i will perform the job with quality $z_{1i} \in [0, \hat{Z}_1]$ as given by a draw from the density of production possibilities $f(z_1)$. If the task is of class Z_2 , a type-1 worker in isolation will not be able to transcribe the questionnaire and will produce 0. Similarly, a type-2 worker j will transcribe the questionnaire with quality $z_{2j} \in [0, \hat{Z}_2]$, drawn from $g(z_2)$ whenever the task is of class Z_2 , and 0 otherwise. Once the job is completed the manager observes each worker’s performance. Figure 1 illustrates the production process.

⁵Higher costs of between-type communication could alternatively be explained by workers’ preference for communicating with same-type colleagues (ingroup) than with others (outgroup).

⁶See appendix 6.1 for a complete list of the main assumptions on which the model rests.

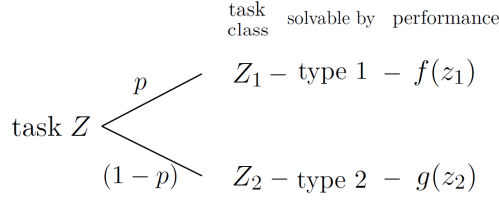


Figure 1 – Task Z in the two different states of the world, the type of worker that can solve it in each case and her conditional performance.

The expected production of workers i and j in isolation is respectively given by $E(z_i) = pE(z_{1i})$ and $E(z_j) = (1 - p)E(z_{2j})$: the probability of the task being of the kind they can solve times the expected value of their performance in this case.

In a group with multiple workers each individual might talk to her colleagues, improving her own performance thanks to their potentially superior knowledge. We assume that group size and communication costs are small enough for each worker to find it optimal to communicate with everyone else. Each worker performs her task as well as the best worker she communicates with, minus the incurred communication costs.

The commitment problem of better workers not wanting to share their superior knowledge once the task and abilities are revealed can be solved in different ways. Better performers would want to help worse ones in a repeated setting, as they could be the lower ability ones in future repetitions of the problem. Another incentive for all workers to communicate their knowledge would be to make individual payment dependent on group performance.

2.2 Knowledge spillovers in homogeneous groups

If worker i is part of a group composed of $m > 1$ workers and communicates with each of her colleagues at a cost c , her expected performance is:

$$E(z_i|m) = pE(z_{1i}|m) + (m - 1)c = pZ_{1m} - (m - 1)c, \quad (1)$$

where Z_{1m} is the expected highest order statistic from a sample of m draws from $f(z_1)$, a concave function of the sample size for any continuous distribution.⁷ The cost of communication c is multiplied by $m - 1$ because worker i is communicating with every other employee in the group except herself.

The concavity of the expected order statistic implies that, for homogeneous groups composed of same-type workers:

Proposition 1. *The effect of group size on expected performance is concave.*

Intuitively, this result is due to the redundancy of same-type workers' knowledge increasing with the number of workers of this type.

⁷See figure 7 in appendix 6.2 for an illustration of the expected order statistic for h draws from a $[0, 1]$ uniform distribution, and De La Cal and Cárcamo 2005 for relevant results related to the concavity of the expected highest order statistic.

In our example, the contribution of a nurse to her unique colleague's expected performance in a group of size 2 is much higher than her contribution to each colleague in a group of size 4. The reason is that the probability of her being significantly better at solving the task at hand than anyone else decreases as group size increases.

The lower contribution of similar workers to the expected performance of their colleagues in larger groups will have a crucial impact on the effect of diversity. In the next section we explore this effect by introducing type-2 workers able to solve task Z_2 into the model.

2.3 Knowledge spillovers in diverse groups

A group of size N is now potentially composed of m nurses (type-1 workers) and k economists (type-2 workers).

A type-1 worker can potentially improve over her expected performance in isolation by communicating both within and between types. To communicate with a worker of a different type costs c_o . For simplicity the cost of within-type communication c is normalized to 0 in this section.⁸

The expected performance of a type-1 worker i in a diverse group with $m - 1$ type-1 and k type-2 colleagues is

$$E(z_i|m, k) = pZ_{1m} + (1 - p)Z_{2k} - kc_o. \quad (2)$$

Her performance in a same-size homogeneous group is

$$E(z_i|m + k, 0) = pZ_{1m+k}. \quad (3)$$

We define the relative contribution $\Delta_k(m)$, of k diverse workers to the expected performance of each of their m type-1 colleagues as the difference between their performance in this diverse group and in a same size homogeneous one:

$$\Delta_k(m) = E(z_i|m, k) - E(z_i|m + k, 0) = p(Z_{1m} - Z_{1m+k}) + (1 - p)Z_{2k} - kc_o. \quad (4)$$

The concavity of the highest order statistic, together with the knowledge of type-2 workers contributing the same in larger and smaller groups, produces the following result:

Proposition 2. *The relative contribution of k diverse workers $\Delta_k(m)$ is increasing in group size m .*

The proof follows directly from the concavity of the expected highest order statistic.⁹

This result is crucial for a proper understanding of the relationship between skills diversity and productivity. Under very general conditions, the same kind of diversity might have a completely different effect in groups that differ only in their size. Importantly, group size does not only affect the dimension of the effect, but it might even change its sign.

As group size increases and the knowledge of type-1 workers becomes redundant, the effect of having a fixed number of diverse colleagues might be positive in relatively large groups,

⁸This is equivalent to interpreting c_o as being equal to the difference between the between-type and within-type cost of communication. Normalizing c to 0 simplifies the model without reducing its explanatory capacity.

⁹Formally, for $l > m$, $\Delta_k(l) - \Delta_k(m) = [E(z_i|l, k) - E(z_i|l + k, 0)] - [E(z_i|m, k) - E(z_i|m + k, 0)] = [pZ_{1l} + (1 - p)Z_{2k} - kc_o - pZ_{1l+k}] - [pZ_{1m} + (1 - p)Z_{2k} - kc_o - pZ_{1m+k}] = p[(Z_{1m+k} - Z_{1m}) - (Z_{1l+k} - Z_{1l})] > 0$.

with $(1 - p)Z_{2k} - kc_o > p(Z_{1l+k} - Z_{1l})$ for $l > m$, while the same effect might be negative in smaller ones, with $(1 - p)Z_{2k} - kc_o < p(Z_{1m+k} - Z_{1m})$.

A numerical example using the performance of the nurses and economists that work transcribing questionnaires should help to illustrate this result. For simplicity and without loss of generality, we compare performance in a homogeneous group and in a same size group with only 1 diverse colleague. There is a relatively high probability of 80% that the relevant task is health related. The conditional ability of both nurses and economists is given by a draw from a uniform distribution with support $[0, 1]$. In this case, the expected performance of a nurse in isolation is $0.8 \times 0.5 = 0.4$. A second nurse in the group contributes to her expected performance by $0.8 \times 0.75 - 0.4 = 0.2$. With between-type communication cost $c_o = 0.01$, having an economist as the only colleague would increase the expected performance of a nurse by $0.2 \times 0.5 - 0.01 = 0.09$. In this case, the expected performance of a nurse is lower if her only colleague is an economist than if she works with another nurse.

In a group composed of 3 nurses, the expected performance for each of them is given by $0.8 \times 0.875 = 0.7$. This increases to $0.8 \times 0.9375 = 0.75$ if they are 4. Hence, the contribution of a fourth nurse is 0.05. Notice that in this case this is significantly smaller than the contribution of an economist, which remains the same at 0.09. In this case, the expected performance of a nurse is higher if she works with two nurses and an economist than if she works with three other nurses.

As shown by this simple example, it is easy to think of situations in which the same kind of diversity has a negative effect in small groups but a positive effect in larger ones. Another example would be the contribution of a sociologist to the performance of a group of computer scientists developing a social media app, relative to that of an extra computer scientist. Two computer scientists would probably be more productive than one of them working only with this sociologist. However, in a sufficiently large group, a sociologist should contribute more to her colleagues' performance than say, a 50th computer scientist.

This analysis shows that whenever production tasks are complex and knowledge spillovers important, it is better for colleagues' performance to place a fixed number of diverse workers in a large group than in a smaller one. In the next two sections we explore the empirical implications of this result.

3 Causal evidence: a lab experiment in Guinea-Bissau.

In the real world, the distribution of employees into the productive units of an organization is rarely a random process. A variety of criteria determining who works where imply that selection issues make it hard to find causal evidence on the relationship between diversity and performance. Our suggested mechanism, requiring also significant heterogeneity in group size, takes this difficulty one step up.

The high cost of artificially replicating this kind of setting in the real world makes a lab experiment optimal for a causal test of the predicted effects. With this aim, we run a lab experiment in Guinea-Bissau. Nursing and Economics university students solve a multiple-choice test twice. First, each participant solves the test in isolation. After this, participants are exogenously sorted into different groups. Each of them solves the same test again, this time being able to discuss their choices with the rest of participants in the same group. Before explaining the experimental set up and the results, we describe the recruitment of the participants and the task.

3.1 Recruitment and task

Recruitment

All the subjects for the experiment were recruited among the students from the 2nd, 3rd and 4th year of the Bachelor's degrees in Economics and Nursing at the Universidade Lusófona da Guiné-Bissau (ULG). The main criterion for the choice of these two degrees was the clear disjointness between the content of their courses. We excluded 1st year students because of the expected lower familiarity with their discipline.

In the announcement, it was stated that each participants would solve some simple tests and would be paid between 500 and 5000 FCFA (between 0.76 and 7.58€) for a 2 hours and thirty minutes session, depending on performance. The payment must have been attractive,¹⁰ as 503 (387 nurses and 116 economists) of the 1104 registered students sign up as available to participate in the experiment.

The required number of participants of each type was randomly chosen among the available students. Given the degree of the students registered, we decided to use nurses as our majority — treated — type and economists as our minority — treatment — one. For each degree, an identical number of female and male students was invited to participate in the experiment.¹¹

After a pilot session with 44 students, we run the main experiment with 258 participants distributed across four sessions in a single day.

The task

Upon arrival, students were welcomed and registered at the reception desk. Once enough participants of each type — as defined by degree and gender — had arrived, participants were randomized into groups, given a personal card indicating their ID for the experiment, and taken to one of the rooms. Once there, the room supervisor would place them far from each other and read out loud the instructions.

Each student would receive 500 FCFA (0.72€) for participating in the study, independently of performance. Each of them had to answer the same test twice and a survey. The test consisted of 25 multiple-choice questions. The first 10 were general-knowledge questions. The next 5 were questions from the other discipline (Economics questions for nurses and Nursing questions for economists). The last ten were questions from the participant's field of study.¹²

First, students had to answer the test individually, without communicating with anyone else, receiving 80 FCFA (0.12€) per right answer, for a possible total of 2000 FCFA (3.03€) per test. Every participant in each session studying the same degree answered the exact same test, and had 30 minutes to complete it. After finishing and handing-in the first

¹⁰In a recent study Alvarez Pereira et al. 2017 estimate about 45% of the population of Guinea-Bissau to live below 2\$ a day. In this country, half of the maximum possible payment in the experiment (2500 FCFA or 3.8€) represents about twice the typical daily wage.

¹¹See appendix 6.3 for a detailed explanation of the recruitment process.

¹²The first five own-field questions for nurses coincided with the five other-field questions for economists, and vice-versa. The last five questions in the test, own-field questions as well, were not included in the other type's test. The general-knowledge questions were suggested by the authors and their colleagues. The discipline-specific questions were designed by professors from the corresponding department of the ULG. Check appendix 6.5 for a more detailed explanation of the kind of questions included

test to their room supervisor, students were organized into groups, as indicated in their randomly assigned card.

The second test was identical to the first one for each participant, and was also to be completed individually. However, this time participants were allowed to communicate and discuss their answers with the other members of the group. As before, each right question was paid 80 FCFA. However, in order to incentivize the best performers to share their knowledge, in this occasion one of the individual tests from the group was randomly chosen and all group-members would be paid according to the number of right questions in this test. To make sure that participants understood the implications, we explicitly included the sentence "*this means that you might be paid according to the quality of the answers given by your group-mate*" in the instructions. Participants' improvement between the individual and the group tests will give us a relevant and comparable measure of the effect of communication on individual performance.

Finally, students completed a survey with personal information, being paid 500 FCFA for it. Everything was paper-based. The tests were corrected on the spot, such that students were paid just a few minutes after finishing the survey. Room supervisors were instructed to place students and groups far enough from each other, such that it was unfeasible for them to cheat.¹³ No incidents were reported.

3.2 Groups and treatments

We designed the experiment to simultaneously study three factors that could affect performance: group size, degree diversity and gender diversity. With this aim, participants were exogenously distributed into groups of size 2 and 4, composed of either nurses or nurses and economists, and either homogeneous or diverse in gender.

To test proposition 1 in the theoretical model we isolated the effect of group size from diversity: part of the nurses were allocated into small and large homogeneous groups composed of same-gender nurses.

We test proposition 2 using two kinds of diversity. First, isolating degree diversity, nurses were placed in groups of size 2 and 4 with one same-gender economist. Second, for the study of gender plus degree diversity, same-gender nurses were placed into groups of size 2 and 4 with one economist of the opposite gender.

The design of the experiment, such that knowledge should be more disjoint across degree than across gender, makes gender a secondary dimension of diversity in this study. Still, the frequent association between gender and discrimination, and its potential impact on communication, makes it valuable to study the effect of gender in detail. We do this by subtracting the effect of degree diversity from the effect of degree plus gender diversity.

Our design also takes care of the possibility that gender has a different effect for female and male participants. Every group has its male and female version, as given by the gender of the nurses that compose it.

Figure 2 shows the different groups in the study. We invited and distributed participants such that each session would have at least one group of each type. The exceeding participants were randomized into groups of size 2. There are a total of 6 large groups of each class, for a total of 36. For each class of small groups, the total number goes from 5 to

¹³A participant would be disqualified without any payment were she to break the rules, including talking to other participants (outside her group) or using her cellphone. In the group test the whole group would be disqualified if a single one of its members were to cheat.

11.¹⁴ All the participants in each session were randomly assigned to their groups at the same time.

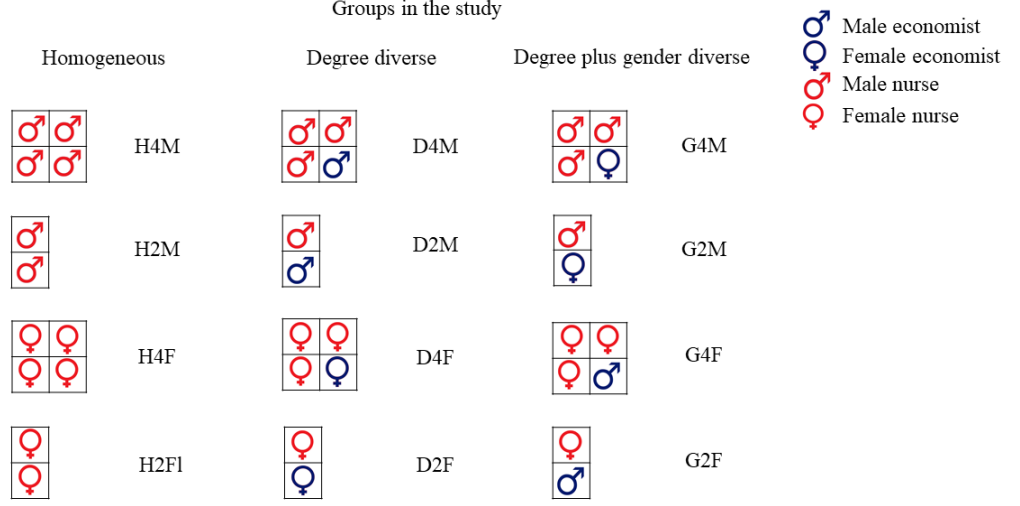


Figure 2 – This figure shows the groups in the experiment. *H* in the name of the group stands for homogeneous, *D* for degree diverse and *G* for gender plus degree diverse. The number, either 2 or 4, indicates the size of the group. The last letter, either *M* or *F*, respectively indicates if the nurses in the group are male or female.

Keeping all these different groups along the analysis of the results is only meaningful if their impact on the outcome of interest is large enough. Our measure of performance is given by each participant’s improvement in the number of correctly answered questions between the individual and group tests, which we will call *gain*. Differences between the median gain for female and male nurses in the same class of group —homogeneous, degree diverse and degree plus gender diverse— are not statistically significant.¹⁵ We collapse the female and male version of each group. Hence, the groups in the study will be just divided along our three dimensions of diversity and their size, see Figure 3.¹⁶

3.3 Checking for balance and descriptive statistics

The exogenous allocation of participants into the previously described groups will allow us to causally test the main predictions from the theoretical model. First, we check for balance across same-type participants assigned to different groups.

A comparison of observable characteristics across same-type participants randomly assigned to different groups shows that they are similar enough: a t-test of differences in the mean number of correct answers in the individual test, and in the value taken by other demographic variables for nurses in small and large groups, is never statistically significant across the three different levels of diversity considered (at the 5%). Only one out of 21

¹⁴There are 5 H2M, 8 H2F, 6D2F, 10 D2M and G2M, and 11 G2F groups.

¹⁵A nonparametric *k*-sample test on the equality of medians shows that none of the pairs of groups compared show statistically significant differences at the 5% and only one of them at the 10%, see appendix 6.4 for details. A parametric t-test on the equality of means produces equivalent results.

¹⁶There are 12 large groups of each class (H4, D4 and G4). For small groups there are 13 H2, 16 D2 and 21 G2.

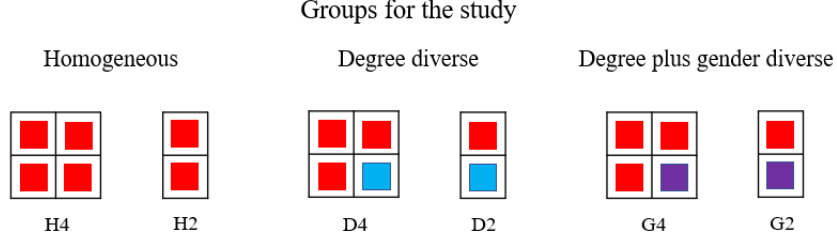


Figure 3 – This figure shows the composition of the groups considered for the analysis of the results from the experiment. *H* in the name of the group stand for homogeneous, *D* for degree diverse and *G* for gender plus degree diverse. The number, either 2 or 4, indicates the size of the group.

comparisons, is statistically significant at the 10%, a finding that can be considered due to chance, see a more detailed analysis in appendix 6.6.

These results show evidence for good balance in pre-treatment outcomes across treatment groups, suggesting that our randomization was successful. Selection should not be a problem for our study.

After checking for pre-treatment balance in our data, we next show descriptive statistics about the different types in our study. We would expect nurses and economists, and probably female and male participants to be different in a variety of characteristics. Table 1 shows the average value for a few variables of interest, organized by type.

	Degree		Nurses		Economists	
	Nurses (1)	Economists (2)	Male (3)	Female (4)	Male (5)	Female (6)
	Number of right questions/ total					
Right answers, all	.48 [.13]	.04** (.02) <i>0.018</i>	.51 [.12]	-.08*** (.02) <i>0.000</i>	.56 [.12]	-.09*** (.03) <i>0.002</i>
Right answers, own-field	.46 [.19]	.01 (.026) <i>0.620</i>	.47 [.19]	-.014 (.093) <i>0.596</i>	.49 [.21]	-.04 (.04) <i>0.333</i>
Right answers, other-field	.21 [.19]	.05* (.03) <i>0.065</i>	.21 [.19]	-.00 (.03) <i>0.694</i>	.28 [.22]	-.02 (.06) <i>0.784</i>
Right answers, general	.62 [.19]	.07** (.03) <i>0.011</i>	.71 [.15]	-.17*** (.024) <i>0.000</i>	.77 [.12]	-.18*** (.04) <i>0.000</i>
Some empty answer (d)	.43 [.50]	-.15** (.06) <i>.021</i>	.42 [.50]	.028 (.07) <i>.694</i>	.27 [.45]	.03 (.11) <i>0.793</i>
Demographics						
Enrollment year	2.60 [.64]	-.031 (.09) <i>0.72</i>	2.76 [.62]	-.31*** (.09) <i>0.001</i>	2.62 [.59]	-.12 (.15) <i>0.423</i>
Age	25.85 [3.62]	-1.05** (.50) <i>0.037</i>	26.16 [3.97]	-.64 (.56) <i>0.254</i>	25.59 [3.70]	-1.70** (.78) <i>0.033</i>
Participants	187	67	92	95	37	30

[STD], (SE), *p-values*, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1 – This table reports and compares average characteristics for nurses and economists, for male and female nurses, and for male and female economists. Columns (1), (3) and (5) show average characteristics of the group of individuals specified by the column heading. Columns (2), (4) and (6) report the difference between the average characteristic of the group in the column to the left and that in the column heading.

The first panel in table 1 shows individual ability in isolation, for different kind of questions. Focusing first on degree diversity, as displayed in columns (1) and (2), we notice that in average an economist answers 4.3% more questions correctly than a nurse. When we analyze the sub-sets of questions, we find that this is mainly explained because economists are significantly better (7%) at answering general-knowledge questions.¹⁷ Another significant difference is that 43% of nurses has left at least one non-answered question, while an economist is 15% less likely to do so. As wrong questions are not penalized, not answering

¹⁷This might be so because general-knowledge questions were designed by economists. Hence, these questions tend more to refer to social, historical and political topics than towards the natural sciences.

a question is obviously sub-optimal.¹⁸ As we can see, for both nurses and economists the proportion of right answers to other-field questions is just above 0.2. This implies that they are just slightly better than guessing at random in answering these questions. Hence, economists and nurses have comparable levels of ability for discipline-specific questions (same-field and other-field), but economists are significantly better at answering general-knowledge ones, what gives them better overall ability.

When we analyze the different levels of ability based on gender we find a similar pattern. For both nurses and economists, females' overall ability is significantly lower than that of their male counterparts (8% lower for female nurses and 9% for female economists). Again, this is explained by a lower ability for answering general-knowledge questions, with female nurses giving 17% less of right answers than male nurses. This difference is even larger between economists of different gender. Interestingly, there is not a single statistically significant difference in the averages for discipline-specific questions (of any kind) across gender. Hence, female participants have a similar-enough ability as their male colleagues at answering discipline-specific questions, and a significantly lower one at answering the general-knowledge ones.

As for the two demographic variables considered in panel 2 of table 1 economists are about one year younger than nurses. Among economists themselves, females are in average 1.7 years younger than males. Female nurses participating in the study were enrolled at a higher year (about 1/3 of a year) than male nurses.

3.4 Results

Communication gains and performance

The participants in the experiment have a diverse background and heterogeneous ability. The payment to the second test, paying everyone in the group according to the performance of a randomly selected member, incentivizes sharing information to help other group-members perform well. Put together, these two ingredients imply that we should see significant gains from communication. As illustrated by Figure 4, this is indeed the case.

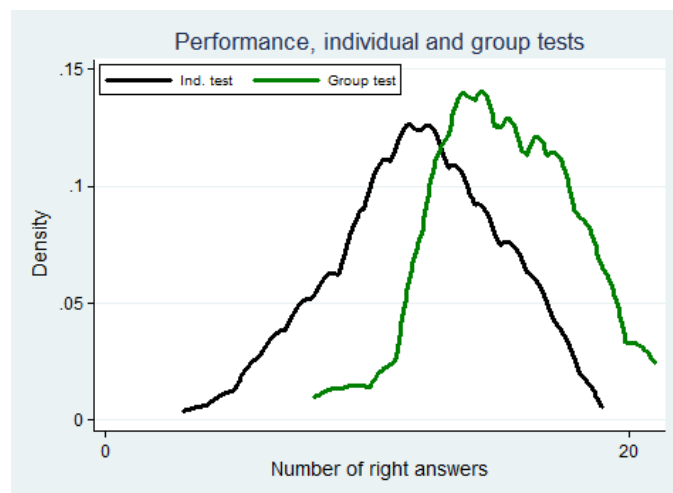


Figure 4 – Distribution of the number of correctly answered questions by nurses.

¹⁸We explicitly wrote in the instructions for supervisors not to comment on this fact, such that the information about leaving non-answered questions was the same for everyone.

The average number of correctly answered questions for nurses is 11.91 (std. 3.13) in the individual test and 15.28 (std. 2.66) in the large one, a 28.3% higher. Equivalently, the average *gain* between the individual and group tests, our measure of performance for the nurses in the experiment, is 3.37 questions (std. 2.93). As we will see, the average gain will be very different for nurses in different groups.

Group size and degree diversity

This experiment offers a simple and clean setting to test the main predictions from the theoretical model:

1. For homogeneous groups composed of same-type participants, agents' performance should be concave in group size.
2. The relative contribution of a diverse participant to the individual performance of her colleagues should be increasing in group size.

Before running any formal tests, the left-hand-side of Figure 5 illustrates the distribution of the number of correct answers gained between the individual and the group tests, for nurses in homogeneous groups of different size. Visual exploration of these two distributions does not help much in understanding which of the two group structures is more favorable for nurses' performance.

The right-hand-side distributions in Figure 5, for the case of degree diverse groups, do point towards a higher average performance for nurses in a large group.

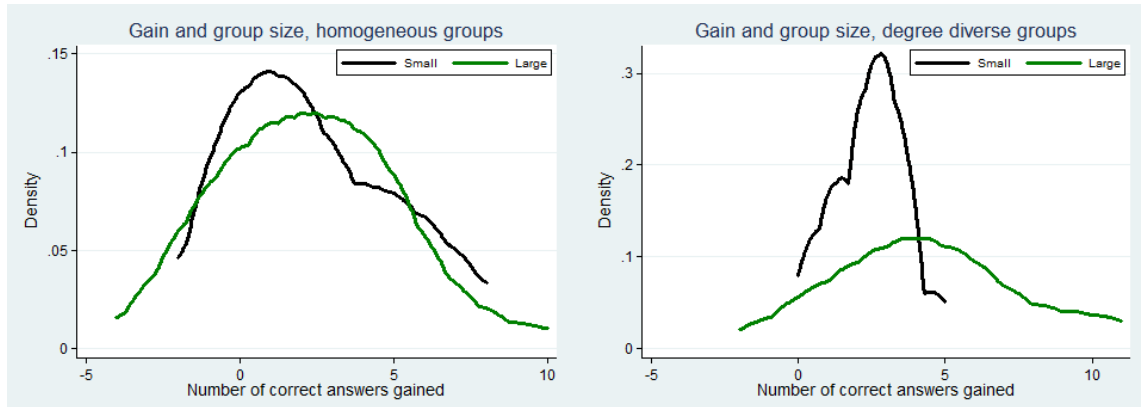


Figure 5 – Distribution of the number of correct answers gained bet the individual to the group test for nurses in small and large groups. LHS: homogeneous groups. RHS: degree diverse groups.

Results 1: performance is increasing in group size at a diminishing rate

Table 2 offers a formal comparison of the performance of nurses in homogeneous groups of different size. A parametric t-test comparing the difference in the average gain for nurses in small and large homogeneous groups does not find this to be statistically significant: nurses do not perform better in large groups than in small ones, when these are homogeneous.

	Homogeneous groups			
	Individual	2 nurses	4 nurses	Diff (2) vs (3)
	(1)	(2)	(3)	(4)
<i>gain</i>	0	2.15	2.25	.09
	-	[2.60]	[2.98]	(.67)
				.886
N	74	26	48	
[STD],(se), <i>p-values</i> , * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$				

Table 2 – This table shows the average value for *gain* for nurses in homogeneous groups of size 1, 2 and 4. Column (4), t-test for the equality between the average *gain* for nurses in groups of size 2 and size 4. This is not statistically different from 0.

Going from a group of size 1 (individual test) to a group of size 2 is associated with an average gain of 2.15 correct answers.¹⁹ Having limited time, the difference in gain between the homogeneous group of size 2 and 4 is not statistically different from 0. This is probably influenced by communication costs: the larger time required for discussing the right answers in the large group compensates for the increase in knowledge associated to the larger number of group members.

These results support the concavity of performance in group size, in homogeneous groups.

Results 2: the relative contribution of a diverse worker is increasing in group size

In the case of diverse groups the average gain is substantially higher in large than in small groups. This holds both for degree diverse and gender plus degree diverse groups.

For degree diverse groups, a nurse paired with a same-gender economist in a small group improves its number of correct answers by 2.44, in average. A nurse paired with another same-gender economist in a large group shows a mean gain of 4.44, 81.9% higher, see table 3.

	Small groups			Large Groups		
	Diverse	Hom.	Diff	Diverse	Hom.	Diff
	(1)	(2)	(3)	(4)	(5)	(6)
Gain	2.44	2.15	.28	4.44	2.25	2.19***
	[1.26]	[2.60]	(.60)	[3.25]	[2.98]	(.70)
			0.639			.002
N	16	26		30	48	
[STD], (SE), <i>p-values</i> , * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$						

Table 3 – This table shows the average gain for nurses in diverse (1) and homogeneous (2) small groups, and in diverse (4) and homogeneous (5) large groups. Columns (3) and (6) show the results for a t-test for the equality of the mean gain respectively for those nurses in columns (1) and (2), and in columns (4) and (5).

¹⁹Using the 5 Nursing questions that economists and nurses do not share in their tests, we can observe that nurses paired with an economist in a small group do not improve the number of correct answers in this sub-set of questions between the individual and the group test. This serves as evidence of all gains coming from communication (and not from repeating the test), allowing us to use the individual test as the size-1 group. Gain is by definition 0 in this case.

Subtracting these effects allows us to compare the relative contribution of a same-gender economist to her colleagues in a large and in a small group. Denoting by $\bar{P}(G)$ the average performance of a nurse in a group of kind G , this is estimated as

$$\left[\bar{P}\left(\begin{array}{|c|c|} \hline \text{■} & \text{■} \\ \hline \text{■} & \text{■} \\ \hline \end{array}\right) - \bar{P}\left(\begin{array}{|c|c|} \hline \text{■} & \text{■} \\ \hline \text{■} & \text{■} \\ \hline \end{array}\right) \right] - \left[\bar{P}\left(\begin{array}{|c|} \hline \text{■} \\ \hline \text{■} \\ \hline \end{array}\right) - \bar{P}\left(\begin{array}{|c|} \hline \text{■} \\ \hline \text{■} \\ \hline \end{array}\right) \right]$$

Formally,

$$\Delta_d(4) - \Delta_d(2) = [\bar{P}(D4) - \bar{P}(H4)] - [\bar{P}(D2) - \bar{P}(H2)] = \mathbf{2.19} - \mathbf{.28} = 1.91, \quad (5)$$

where $\Delta_d(m)$ stands for the relative contribution of a degree diverse worker to nurses in a group of size m . A nurse paired with an economist in a small group gains 2.44 correct answers through communication, 0.28 more than when paired with a nurse, or a 13% more, and this difference is not statistically significant: it is not better for a nurse to be paired with a single economist than to be paired with one nurse. A nurse paired with an economist in a large group gains 4.44 correct answers, compared to 2.25 when placed with other three nurses, a 97.3% higher gain.²⁰ Figure 6 illustrates the size of these effects.

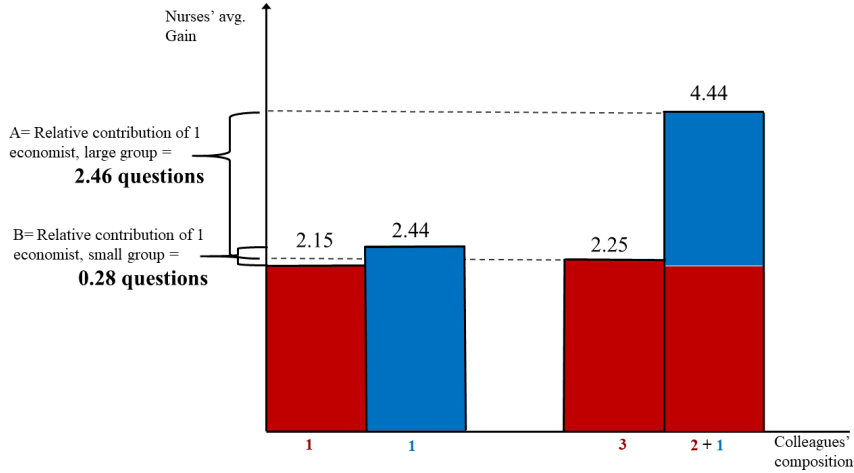


Figure 6 – Relative contribution of a same-gender economist in a large and a small group.

This result supports proposition 1 from the theoretical model: the relative contribution of a diverse worker is increasing in group size.

While the relative contribution of an economist to the only colleague in a small group is basically 0, having an economist instead of a fourth nurse in a large group almost doubles the gains from communication for her three colleagues. Interestingly, this effect holds for every sub-set of questions in the study, see appendix 6.7

After this result for degree diverse groups, we would expect proposition 2 to hold also for degree plus gender diverse groups. With knowledge being more disjoint across degree than across gender (see panel 1 in table 1), we expect the effect of gender to be dominated by that of degree diversity. Indeed, visual examination of the distribution of the gain for nurses in small and large gender plus degree diverse groups shows that this is similar to the distribution in degree diverse groups, see Figure 9 in appendix 6.8.

²⁰This represents a 73% of the standard deviation of gain. Interestingly, economists themselves perform better in degree diverse large groups than in small ones (gain = 3.5 vs 2.20), though this difference does not reach statistical significance.

As for the case of degree diverse groups, we test proposition 2 by comparing the average performance of nurses in gender plus degree groups with that of nurses in homogeneous groups. Table 4 shows gain for nurses in small and large gender plus degree diverse groups.

Gender plus degree diverse groups			
	Small (1)	Large (2)	Difference (3)
Gain	3.48 [2.06]	4.94 [2.71]	1.47** (.64) .026
N	21	36	
[STD], (SE), <i>p-values</i> , * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$			

Table 4 – This table reports and compares the average gain in the number of correctly answered questions by nurses in gender plus degree diverse groups. Columns (1) and (2) respectively show the average gain for those nurses in small and large groups. Column (3) reports the results of a t-test for the equality of means.

For the comparison with homogeneous groups we use the results from table 4 and those from columns (1) and (2) in table 2. The difference in the relative contribution of a gender plus degree diverse participant to her colleagues, between a large and a small group is estimated as:

$$\left[\overline{P}(\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}) - \overline{P}(\begin{smallmatrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{smallmatrix}) \right] - \left[\overline{P}(\begin{smallmatrix} \blacksquare \\ \blacksquare \end{smallmatrix}) - \overline{P}(\begin{smallmatrix} \blacksquare \\ \blacksquare \end{smallmatrix}) \right]$$

Formally,

$$\Delta_{dg}(4) - \Delta_{dg}(2) = [\overline{P}(G4) - \overline{P}(H4)] - [\overline{P}(G2) - \overline{P}(H2)] = (4.94 - 2.25) - (3.48 - 2.15) = 1.36, \quad (6)$$

where $\Delta_{dg}(m)$ stands for the relative contribution of a degree plus gender diverse worker to nurses in a group of size m , as compared to their performance in a homogeneous group of size m . The relative contribution of a diverse worker in a large group (2.69 questions) is still much larger than her relative contribution in a small one (1.33 questions).

This section shows that results for both degree and gender plus degree diverse groups satisfy proposition 2 in the theoretical model: the relative contribution of a diverse worker to colleagues' performance is larger in larger groups.

Group size and gender diversity

As we have seen in the description of the groups, gender was introduced as an extra dimension of diversity in part of the degree diverse groups, instead of introducing it directly into degree homogeneous ones.²¹ Hence, a comparison of these two classes of groups will allow us to isolate the gender effect.

²¹This choice was based on our preference for having more observations for degree diverse groups, in case empirical findings implied that these could be merged with degree plus gender diverse groups, in case there was not a significant gender effect.

The first we see when we compare the effect of gender plus degree diversity with that of degree diversity is that, for same size groups, more diversity is associated with a higher performance. The effect of gender diversity is identified by subtracting the effect of degree diversity from the combined effect, by doing:

$$\left[\overline{P}(\text{Gender+Degree}) - \overline{P}(\text{Degree}) \right] - \left[\overline{P}(\text{Gender}) - \overline{P}(\text{Control}) \right]$$

Formally,

$$\Delta_g(4) - \Delta_g(2) = [\overline{P}(G4) - \overline{P}(D4)] - [\overline{P}(G2) - \overline{P}(D2)] = (4.94 - 4.44) - (3.48 - 2.44) = -0.55, \quad (7)$$

where $\Delta_g(m) = \Delta_{gd}(m) - \Delta_d(m)$ gives the relative contribution of a gender diverse worker to nurses in a group of size m . This is equivalently estimated as the difference between the contribution of a gender plus degree diverse worker minus that of a degree diverse one: $[\Delta_{gd}(4) - \Delta_{gd}(2)] - [\Delta_d(4) - \Delta_d(2)] = 1.36 - 1.91$, resulting again in a gain equal to -0.55.

Gender diversity has a larger impact in smaller groups than in larger ones.

The knowledge of same-gender participants should become more redundant as their number increases than that of the one participant of a different gender. Hence, these findings seem to point towards the existence of some case of discrimination of colleagues of a different gender when these are a minority within the group.

As we have seen, the non-parametric tests of the median gain across different gender groups with the same kind of diversity did not find statistically significant differences across them. However, a comparison of the mean gain for participants of different gender in gender plus degree diverse groups offers a clear pattern, that strongly supports the possibility of this result being caused by discrimination: the gender effect is larger for male participants in larger groups, and smaller for female participants:

- For female nurses: $[\Delta_{gd}^f(4) - \Delta_{gd}^f(2)] - [\Delta_d^f(4) - \Delta_d^f(2)] = (5.05 - 4) - (5.11 - 2.5) = -1.56$.
- For male nurses: $[\Delta_{gd}^m(4) - \Delta_{gd}^m(2)] - [\Delta_d^m(4) - \Delta_d^m(2)] = (4.83 - 2.9) - (3.78 - 2.4) = 0.55$.

This is also true for female and male economists.²²

This result is closer in nature to the identity side of diversity. Differences in preferences for the utility of ingroup and outgroup members would determine participants behavior toward workers of a different type. In the model, this could be explained by the introduction of endogenous costs for between-gender communication that increase as the ratio between the number of workers of different types in the group gets away from parity, possibly taking into account non-symmetric discrimination between genders.

The fact that participants discriminate gender diverse group members even when the formers' payoff is directly dependent on the latter's performance is particularly striking. It seems difficult to find a rational explanation for this behavior.

This result shows that a larger group size might increase the negative effect of discrimination on performance. This effect goes in the opposite direction of the standard effect of group size, when diversity in knowledge exists without discrimination. Hence, before

²²In this case: for male economists the difference is 0.67, while for female economists this is equal to -4.

thinking about the effect that diversity should have in small and larger groups, we should carefully consider the extent to which the context and the kind of diversity present might favour discrimination.

This experiment has allowed us to find causal support for the main predictions from the theoretical model in a controlled environment. In the next section we support these findings with stylized facts from the empirical study of the performance of Pakistani tax collectors.

4 Collecting taxes in Punjab

In this section we explore the role of group size, skills diversity and individual performance in the context of an organization that fits the model’s set-up and assumptions. Civilian and ex-military civil servants are responsible for tax collection in municipalities that are grouped into districts of different sizes, with a common supervisor and monthly meetings that offer the potential for relevant knowledge spillovers. The repeated nature of interactions and the large variety of tasks a worker needs to solve should create the right incentives for workers to share useful information among themselves. In this setting, workers of two clearly differentiated types with relevant and heterogeneous skills solve an individual complex task with a comparable measure of performance, and there is a coherent definition of groups of different sizes. As we will see, while selection issues will not allow for a causal identification of the effects, our findings go in the same direction as the results from the experiment, supporting the predictions from the theoretical model.

We first describe the main elements of the setting. This should help the reader to understand how the main elements from the theoretical model match this organizational setting, such as how knowledge spillovers might matter, or the motivation behind the definition of workers’ types and groups.

4.1 The organization, the subjects and diversity

With over 110 million of inhabitants by 2017, the province of Punjab contains more than half of the total population in Pakistan. This province is divided into 36 districts, each sub-divided into a number of teshils²³ ranging between 2 and 7, for a total of 141. The public administration has an important presence at these sub-regional levels, being in charge of land tax revenue, among other tasks.

The subjects in this study: elite civil servants

The subjects in this study are civilian and a minority of ex-military elite civil servants working in the districts and tehsils of the Pakistani province of Punjab.

These civil servants are the highest-ranked workers in the public administration. Civilian civil servants are selected through highly competitive exams, while those coming from

²³These teshils are somehow the equivalent of —large— European municipalities or councils.

the army are selected in an informal way among a pool of military applicants.²⁴ Once they become part of the public administration, these workers occupy a large variety of positions and are responsible for different tasks and objectives in several departments, including revenue, district level finance, education and health.²⁵

Importantly, those civil servants working in the same district take part in periodic meetings, typically on a monthly basis, to discuss their objectives, challenges and performance. These meetings imply that elite civil servants in the same district know and talk to each other frequently, creating the opportunity for relevant knowledge spillovers.

We have data on a total of 1372 civil servants between February of 1987 and December 2013.

Introducing diversity: the ex-military civil servants

Since 1980 a quota of 10% of the positions in the Pakistan Administrative Services are officially allocated to members of the Pakistan Armed Forces. Those members of the military who join the civil service are chosen from a pool of voluntary applicants, without the requirement of taking the official exam, and no clear rules. Once they join the civil services they retire from the army. 70 out of the 421 PAS workers²⁶ in our data joined the civil service in this way.

Having both civilian and ex-military workers is a particularly well suited condition for the study of diversity in an organizational setting. A large number of studies show that self-selection, training and professional experience make the average military person significantly different from the average civilian, typically exhibiting higher professionalism, altruism, authoritarianism and conservatism —see for example Bachman et al. (1987). Akerlof and Kranton (2005) put forward solid arguments and references supporting that military and civilian workers are different in aspects that should make a significant difference in any organizational setting. Importantly, as shown by Jackson et al. (2012), the changes in personality traits experienced by individuals trained in the army have long-lasting effects, persisting for years after individuals have entered colleague or the labor market.

Some of these aspects should be relevant for professional performance and interpersonal communication, such as personality traits, their understanding of professional identity and organizational culture, and consequently their behavior at work.²⁷

It should be clear by now that one should expect the average civilian and the average ex-military worker to diverge in several unobservable aspects. In our data, they are also different in many observable aspects: ex-military workers join the civil services being older, are younger in average, have been promoted more times, are men, and a much higher proportion are original from other regions, see table 5.

²⁴Civilian workers belong either to the Provincial Management Services (PMS) —created in 2004 from the union of the Provincial Secretariat Services and the Provincial Civil Service— and who can only work in the province of Punjab, or the Pakistan Administrative Services (PAS) —selected at the national level—who can work in any province. Those workers coming from the army are part of the PAS. This is the only difference between the PMS and PAS that is relevant for this study.

²⁵They supervise and are assisted by a large number of lower-ranked workers for which there is no data available.

²⁶As said before, ex-military civil servants join the PAS. In our data 817 civil servants belong to the PMS and 421 to the PAS.

²⁷The direct relationship between personality traits and consequential outcomes is a well established fact —see Ozer and Benet-Martinez 2006 for a broad review on the literature on this precise question.

	Ex-military workers	Civilian workers
Joining age (months)	383.25	353.59
Current age (months)	465	497
Women	0	0.07
No Punjabi	0.47	0.13
Number of promotions	1.09	0.30

Table 5 – Average values for different variables across the ex-military and civilian civil servants.

The institutional structure of the Pakistan army, fully based on volunteer enlistment, creates a convenient clear-cut distinction for the military experience of the civilian and the ex-military workers in our sample: all those who have served in the army chose to do so, and virtually none of the civilians would have ever undertaken any military training or had professional military experience. This allows us to divide our sample into two different types of civil servants: civilians will be the type-1 workers, and the ex-military civil servants the type-2 ones.

The differences in observable characteristics, together with the unobservable attributes we would expect to exist between them, should be enough for us to consider the civilian and the ex-military as two different categories of workers. Naturally, the best test will be if the way they are combined into groups affects performance in any measurable way.

4.2 The task and our measure of performance

Among the different positions that elite civil servants might occupy in the district, one of them is of particular interest for this study. Assistant Commissioners (ACs) are grade-17 officials that represent the maximum authority in relation to the "general management, administration of the State land, revenue matters, and coordination between the government departments" in the tehsil where they are allocated.

Over the years, many different legal and civil services reforms have resulted in a change in the role played by ACs (see appendix 6.9 for a description of the main functions of ACs in the province of Punjab under the different regulatory regimes). Importantly, across all reforms, revenue administration has always remained a core duty for Assistant Commissioners.

The task: managing tax collection

The main responsibilities of an AC with regards to revenue administration are to recruit the lower revenue field staff, to monitor their progress, periodically checking their revenue record, and to report the outcome to higher authorities. Hence, in practice every Assistant Commissioner is the direct manager of a relatively large group of revenue officers whose main duties are land administration and collecting taxes from agricultural producers.²⁸

The successful solution of the tasks associated to his position requires the AC to master a diverse set of skills. For an AC to successfully coordinate a team of revenue officers

²⁸The 'Patwari' or 'village accountant' is the official at the lowest rung of the revenue collection machinery. His job encompasses visiting agricultural lands and maintaining a record of ownership and tilling, with their jurisdiction typically spanning a couple of villages. The 'Naib-Tehsildar' is the supervisory tier for a group of 'Patwaris'. These are typically supervised by one or two 'Tehsildars', below the direct supervision of the corresponding AC.

requires, among others: having a broad vision and understanding of the overall system he is supervising, assessing the need for qualified personnel to fill capacity gaps, a relatively good knowledge of the territory, keeping a right monitoring system, promoting an organizational culture which encourages values and ideas that contribute towards a good performance, making sure that his supervisees receive structured and clear mandates, and staying updated about legal and regulatory changes, or technical developments —see Enemark and Van der Molen 2008 for a systematic analysis of the main factors relating workers capacity and an effective land administration.

On top of the long list of complex tasks and skills required from a successful Assistant Commissioner, some institutional particularities of tax collection in Pakistan, such as the corruption of lower revenue officials, complicate their work even further.²⁹

We only have data on the elite civil servants in the district, without any information on the lower revenue officers. A relevant question is whether the skills, knowledge and behavior of the Assistant Commissioner can make any difference to the aggregate tax collected in the tehsil he administrates. In a similar context, Bertrand et al. (2016) finds that Indian elite civil servants who enter the public administration older and in larger cohorts are less effective, and this translates into a lower aggregate economic performance in the states where they are placed.

In this study we explore the existence of systematic correlations between ACs' characteristics and conditions and the amount of taxes collected in the tehsil they administrate.

The measure of performance.

The administrative data collected by the Pakistani Administrative Services contains a quantitative measure of the amount of taxes collected in a given tehsil in a given month, over the centrally established annual objective. This gives us the monthly performance measure for the ACs:

$$\text{Tax performance} = \frac{\text{Tax collected in the month}}{\text{Annual tax target.}}$$

This performance measure is compiled from historical records of the provincial Board of Revenue (BOR) on Land Revenue/Agriculture Income Tax (AIT) from 1983-2013. The Punjab Agricultural Income Tax (AIT) Act 1997 replaced the Punjab Land Revenue Act, 1967. Land Revenue contemplated a tax on area cultivated, while AIT was meant to impose a tax on the agricultural income of an owner in a tax year.

The high difficulty in the administrative tracking and documentation of agricultural income in the rural economy of Pakistan, together with the fact that the revenue administration is obliged by law to keep a detailed record of land owned, imply that for all practical purposes agricultural taxes have remained a tax on area sown (Nasim, 2012). The only of

²⁹The position of 'Patwaris' as the maximum authority for land titling and their important role in tax collection in rural areas makes them particularly prone to corruption, with practices such as demanding bribes and changing land records at will being considered endemic to the post. The perception of widespread corruption and inefficiency among Punjab 'Patwaris' is studied in a series of publications related to the Land Record Management Information System (LRMIS) project for the computerization of land records in this province, see Deininger et al. (2001); Qazi (2005, 2006). The data in our study correspond to the period before computerized land records became functional.

the two rules for tax collection contemplated by the AIT³⁰ which is enforceable in practice is the one for which the revenue authorities keep a detailed record i.e. farmland owned. Hence, in practice the BOR sets annual tax collection targets based on the official record of size and number of farmlands.

Given the centrality of revenue administration among an AC responsibilities, this measure of performance should be an adequate indicator of the degree to which a given AC is doing ‘a good job’.

Before explaining and exploring the data we offer a careful definition of our definition of the groups in this context, and explain their composition

4.3 The definition of groups: elite civil servants in the same district

As mentioned before, each of the 141 tehsils in Punjab is an administrative sub-division of one of its 36 provincial districts. Assistant Commissioners have frequent meetings with those elite civil servants working in their same district at the same time. For our definition of groups, we care about whether other civil servants might influence ACs’ performance. Let us reflect for a moment if this should be the case.

On the one hand, the Assistant Commissioner is the maximum authority for tax collection in the tehsil where she is posted, and manages her group of revenue officers pretty independently. High-powered incentives are relatively scarce in this public sector setting, as civil servants cannot be fired, and salary depends to a large extent on seniority, reducing to a certain extent the pressure or control that other colleagues, such as supervisors, might place on the work of an AC.³¹ This autonomy of action supports skepticism towards the degree to which any colleagues apart from those supervised by the AC —on whom we do not have data—might influence his performance.

On the other hand, we have seen that the Assistant Commissioner has to solve a huge number of complex tasks in order to fulfill her daily responsibilities. Can she know the optimal solution to each of the challenges she faces? Probably not. This makes those elite civil servants with whom the AC has a frequent contact a possible asset in terms of knowledge and skills, as one would expect each of them to have some expertise on at least part of the issues the AC might struggle with.³²

One can think about the taxes collected in a given tehsil in a certain month as the output of a production function corresponding to a process managed by the AC. The information content of this function opens the space for colleagues who do not directly work on tax collection in the same tehsil, but with whom the AC communicates frequently, to have an impact on her performance through knowledge spillovers, potentially influencing the amount of tax she collects in a given period.³³

³⁰See appendix 6.10 for an official description of the two rules contemplated for setting the agricultural tax corresponding to a given producer, as regulated by the AIT.

³¹Promotions is an aspect which should be explored in depth, as a possible means of influence on an AC’s behavior.

³²See for example Rodan and Galunic 2004 for evidence on the positive association between managers’ access to knowledge heterogeneity and their individual performance.

³³Some of examples of relevant and transferable knowledge might include a higher familiarity with the territory or the population in the AC’s tehsil, personally knowing some of the AC’s supervisees, land-owners or other people relevant for tax collection in the community; having faced a challenge similar to that the AC is currently facing, for example previous experience with complicated situations related to the management of human resources, such as incentivizing a un-motivated colleague, or dealing with a supervisee who is suspected corrupt; or expertise with developing good organizational practices such as scheduling functional and effective meetings, or improving the planning of the main objectives to be

The autonomy and authority of the AC in those issues related with the administration of her tehsil implies that she has always the last say in choosing the action she considers better suited for solving the task at hand. This is important for two reasons. First, because a group is defined in this context by social, geographical and professional proximity—which arguably results in ease of communication—and not by a common objective or task, as the Assistant Commissioner’s objectives remain mainly as her personal competency. This has the advantage of making a long discussion about the nature of the different tasks the group faces and their relative relevance unnecessary.³⁴ Secondly, because as long as we assume that ACs are better than random at choosing the best option when presented with several alternative solutions, we can just assume that having more information is always weakly better for an AC’s performance than having less information. This, together with ACs preferring to do better than worse for the same amount of effort, is all we need for our suggested mechanism to work.

Following the criterion of a high enough expected communication frequency among group members, an AC’s group is defined as composed by all those elite civil servants working at the same district at the same time.

We will use this simplest definition of a group for the analysis, considering only the number and type of colleagues. There are 2993 groups in our dataset for which at least a civilian AC has information on performance. These add up to 6935 observations.

4.4 Empirics

The empirical study is developed using a novel panel dataset in monthly frequency, built from confidential documents belonging to the Pakistan Administrative Services—see appendix 6.11.1 for a detailed description of the origin of these data. In this section, we first describe the data used for the study of the role of group composition in explaining the individual performance of Pakistani tax collectors. After describing the available data on Assistant Commissioners’ performance, targets for tax collection and the distribution of types, we proceed to explore the main predictions from the theoretical model.

An observation in our data corresponds to a civilian AC in a given month. There are a total of 6935 observations in our dataset. The average monthly performance across all civilian ACs is 7.69% of the annual target, with a standard deviation equal to 13.65%—the minimum value is 0% and the maximum 100%. As one would somehow expect, average performance is typically higher for those months closer to the end of the financial year (December to June), and lower for the first five months in the financial year (July to November), see figure 14 in appendix 6.12.

As explained in section 4.2, centrally established tax targets for each district depend on the area dedicated to agricultural land. A regression of current tax targets on tax collected in the previous year, agricultural and irrigated area and other controls show that these are independent of previous performance,—see table 12 in appendix 6.11.

We have data on a total of 1372 civil servants between February of 1987 and December 2013. Among these workers, 649 of them work at some point during these time interval as Assistant Commissioners (ACs). Other 723 colleagues work only in other positions (not as ACs), in districts and months in which there is information on performance for

completed during the corresponding financial year.

³⁴For a classification of kinds of tasks in group environments, in categories such as disjunctive or conjunctive, intellectual, criterion or judgmental, maximizing, etc, see for example (Steiner, 2007).

at least one AC. Table 6 shows the distribution of workers across different types, and the corresponding number of observations.³⁵

	ACs, with info on performance		Other civil servants	
	Workers	Observations	Workers	Appearances
Civilian	624	6938	1302	29115
Ex-military	25	170	70	920
Total	649	7108	1372	30035

Table 6 – Number of different ACs of each type and the corresponding number of observations. Other workers and the number of times they appear as part of a group in the dataset.

4.4.1 The main variables

The variables created from the official documents will allow us to study the relationship between group size, diversity and performance. The main variables used in the study are the following:

- *performance*: the performance of tax collector (AC) in the corresponding period, given by *tax collected/annual tax target*.
- *colleagues*: the number of colleagues for the AC, in a given district and period (it excludes the AC himself). The average of this variable is 11.57 (std. 13.60), with the minimum being 1 and the maximum 115.
- *colleagues*⁴: the variable *colleagues* exponentiated, in order to capture the hypothesized non-linear effect of the number of colleagues. After exploring different exponents, we choose to raise the variable to the power 4, as this value seems to maximize the significance of the two coefficients associated to *colleagues* and its exponential version.³⁶
- *military colleagues*: the number of ex-military colleagues in the current district and period.
- Controls: these include worker time-varying characteristics: a dummy variable which takes value 1 if the worker is in its first ever post and 0 otherwise and the age of the worker in months.
- Fixed Effects: in most regressions we will include Fixed Effects for the individual worker, the current district, financial year (July of one year to June of the next one) and month (January to December).

Table 7 shows descriptive statistics for the value these variables take for civilian ACs.

Variable	Obs.	Mean	Std. Dev	Min	Max
<i>performance</i>	6935	7.69%	13.65%	0	100
<i>colleagues</i>	6935	11.57	13.60	1	115
<i>military colleagues</i>	6935	0.28	0.59	0	5

Table 7 – Simple descriptive statistics for the main variables in the study.

³⁵Among the civilian civil servants, 211 ACs and 421 workers who do other jobs come from the PAS. Notice the number of ex-military workers is above the official quota of a 10% of the PAS.

³⁶Other exponential transformations of *colleagues*, such as taking its squared value, offered qualitatively equivalent results with lower statistical significance.

We use these variables to test the predictions from the theoretical model.

4.4.2 Testing the theory: stylized facts

A new member of the PMS or the PAS typically becomes an AC after completing his one-year training. There is not a clear set of rules or criteria for his allocation into one or another tehsil, and so this process cannot be considered random. Potential selection issues imply that these empirical results do not attempt to be a causal explanation of the suggested mechanism, but to serve as evidence supporting the main predictions from the theoretical model.

The analysis is developed using pooled Ordinary Least Square (OLS) regressions. We first explore the relationship between group size and performance in homogeneous groups composed only by civilian workers. After that we describe our empirical approach to analyze the relation between having ex-military colleagues in the group and civilian ACs' performance.

Results 1: performance is increasing in group size at a diminishing rate

This first part of the empirical analysis is developed focusing on the number of colleagues, those elite civil servants working in the same district as the corresponding Assistant Commissioner. The analysis is developed considering only those homogeneous groups that do not have any ex-military worker in that given month.

There are 2302 homogeneous groups in our data, with an average size of 9.12 workers (std. 9.67), going from 2 to 116. Each group has in average 2.34 ACs with information on performance (std. 1.36), from 1 to 7. This gives 5428 observations for ACs in homogeneous groups, with average performance being 7.45% (std. 13.53%), taking values from 0% to 100%.

The theoretical model predicts a larger number of colleagues in a homogeneous group to be associated with a higher performance at a decreasing rate. Proposition (1) implies that we should expect the performance of a worker in a given group or district to be higher when the number of workers in this district is larger.

We start from a linear equation that estimates the *performance* of tax collector i in tehsil k and district d , in month m and financial year t , as:

$$performance_{ikdmt} = \alpha_{i1} + \beta_1 colleagues_{dmt} + \delta_1 X_{ikdmt} + \varphi_{d1} + \lambda_{m1} + \phi_{t1} + \epsilon_{ikdmt}, \quad (8)$$

where β_1 is our main variable of interest, indicating the effect of an extra-colleague on worker i 's performance. α_{i1} is the fixed effect for individual i , X_{ikdmt} are tax collector i 's time-varying observable characteristics, and φ_{d1} , λ_{m1} and ϕ_{t1} are respectively district, month and financial year fixed effects. The error term, ϵ_{ikdmt} is clustered at the tehsil level.³⁷

For the data to support proposition 1 the coefficient β_1 should be positive. As the results in column (1) of table 8 show, in a regression without a non-linear term this coefficient is indeed positive and marginally significant (at the 10%).

According to the theory, a non-linear measure of the number of colleagues is a potentially important element for the right assessment of the association between the number of

³⁷For the remaining of the paper the error term will be clustered at the tehsil level in every regression.

group members and individual performance. We incorporate this non-linear element to regression (9) by including *colleagues*⁴ as an extra variable:

$$performance_{ikdmt} = \alpha_{i2} + \beta_{2a}colleagues_{idmt} + \beta_{2b}colleagues^4 + \delta_2X_{ikdmt} + \varphi_{d2} + \lambda_{m2} + \phi_{t2} + \epsilon_{ikdmt} \quad (9)$$

Proposition (1) in the theoretical model would be supported by the data if $\beta_{2a} > 0$, and $\beta_{2b} < 0$. The results for regression (9), in column (2) of table 8 show that both conditions are satisfied: β_{2a} is positive and β_{2b} is negative, and both are significant at the 1%.

While the magnitude of β_{2b} might seem very small, together with β_{2a} this coefficient implies a coherent interpretation of the marginal effect of an extra colleague on the performance of civilian tax collectors. While an extra colleague in small groups is associated to an increase of the 0.18% of the annual target for an ACs' average performance, that same contribution becomes negligible for large groups with over 75 workers.³⁸

The contribution of civilian colleagues to an ACs performance in their same group is very large in quantitative terms. Everything else equal, going from 5 to 25 colleagues is associated with an increase in average performance of approximately 4.64% of the annual target. This is over 60% of the average monthly performance for civilian ACs, and about one third of its standard deviation.

	performance (1)	performance (2)
colleagues	0.0312* (0.0162)	0.241*** (0.0715)
colleagues ⁴		-1.06e-07*** (3.06e-08)
Controls	Yes	Yes
District FEs	Yes	Yes
Financial year FEs	Yes	Yes
Month FEs	Yes	Yes
Individual FEs	Yes	Yes
Observations	5428	5428
R^2	0.245	0.246
Standard errors in parentheses		
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$		

Table 8 – Pooled OLS for the study of group size in homogeneous groups populated only by civilian civil servants.

In summary, as predicted by the theoretical model, having an extra colleague is associated with a higher individual performance for civilian tax collectors in Punjab, and this marginal gain decreases with the number of colleagues in the group.

³⁸See appendix 6.13 for an illustration of the interpretation of this marginal effect of an extra colleague, conditional on group size.

Results 2: the relative contribution of a diverse worker is increasing in group size

In this section we proceed to explore the relationship between working with ex-military colleagues in the same district and the individual performance of civilian ACs.

Proposition (2) predicts that having a fixed number of ex-military colleagues should be associated with a higher performance for civilian tax collectors in larger groups than in smaller ones.

There are 691 diverse groups in our data, in which at least a civilian AC works with at least an ex-military colleague. The number of ex-military in diverse groups falls between 1 and 5, with an average of 1.268 —see appendix 6.14 for the distribution. These groups are larger in average than the homogeneous ones, with the mean number of workers being 12.70 (std. 11.03), and taking values between 2 and 64. There are 1507 observations for civilian ACs in these groups, with the average performance being 8.52% (std. 14.03%), also larger than in homogeneous groups.³⁹

Including the variable *military colleagues* into regression (9) with the extended sample including also diverse groups, shows a positive coefficient, but this is not statistically significant —see table column (1) in table 9. In a second regression, we include also an interaction between this variable and *colleagues*:

$$performance_{ikdmt} = \alpha_{i3} + \beta_{3a}(military\ colleagues \times colleagues)_{idmt} + \gamma_{3a}colleagues_{idmt} + \gamma_{3b}military\ colleagues_{idmt} + \gamma_{3c}colleagues^4_{idmt}X_{ikdmt} + \varphi_{d3} + \lambda_{m3} + \phi_{t3} + \epsilon_{ikdmt}. \quad (10)$$

The result for this regression is shown in column (2), table 9. The main parameter of interest, β_{3a} is positive and statistically significant (at the 10%, p-value 0.053). This piece of evidence is consistent with proposition 2: the relative contribution of an ex-military worker is increasing in group size.

Interestingly, results are also coherent with another direct implication of the model: the diminishing contribution of ex-military colleagues to civilian ACs' performance as their number increases.⁴⁰ The introduction of the variable *military colleagues squared* into regression (10) gets a negative and statistically significant coefficient (at the 1%). Notice that the coefficient associated to *military colleagues* becomes larger and also significant at the 1% when its exponential version is included in the regression. The marginal effect of an extra ex-military worker on the performance of his AC colleagues is given by:

$$\frac{\partial performance}{\partial military\ colleagues} = 0.085 \times colleagues - 2 \times 0.656 \times military\ colleagues. \quad (11)$$

For a civilian tax collector, having a first ex-military colleague is associated with an increase in individual performance equivalent to 0.85% of the annual tax target in a district with 10 workers, and to an increase of 2.125% in a district with 25. A second ex-military would be associated to a (0.656%) lower increase: 0.194% in the first case, and 1.47% in the second one.

These results for the performance of civilian tax collectors, both in homogeneous and diverse groups, support the main predictions from the theoretical model. Their average performance is increasing in group size at a diminishing rate. Having an ex-military colleague is associated to a higher average performance for larger groups, and this relationship is weaker as the number of ex-military colleagues increases.

³⁹Homogeneous and diverse groups add up to a total of 2993, and to 6935 observations for civilian ACs.

⁴⁰This is equivalent to proposition 1 for type-2 workers.

	performance (1)	performance (2)	performance (3)
military colleagues	0.434 (0.685)	-0.735 (0.746)	0.0539 (0.865)
military colleagues $\times$ colleagues		0.0605* (0.0310)	0.0849*** (0.0310)
military colleagues squared			-0.656*** (0.247)
colleagues	0.145** (0.0681)	0.150** (0.0722)	0.152** (0.0750)
colleagues ⁴	-7.36e-08** (2.96e-08)	-7.18e-08** (3.17e-08)	-7.16e-08** (3.30e-08)
Controls	Yes	Yes	Yes
District FEs	Yes	Yes	Yes
Financial year FEs	Yes	Yes	Yes
Month FEs	Yes	Yes	Yes
Individual FEs	Yes	Yes	Yes
Observations	6935	6935	6935
R^2	0.224	0.225	0.225

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9 – Pooled OLS for the study of the effect of working with ex-military colleagues, as compared to working in homogeneous groups.

5 Conclusion

Exploring Lazear’s model for the individual performance of workers in groups of different sizes has produced two simple but powerful predictions, relevant for the efficient allocation of workers in contexts in which the solution of a complex task makes knowledge spillovers important for production, and employees with skills that are both diverse and relevant are available.

First, in a group populated by similar workers, group size has a concave effect on performance, with the contribution of an extra colleague decreasing as their number increases. Second, as a direct implication of this effect, the relative contribution of a fixed number of diverse workers to the contribution of their colleagues’ performance is increasing in group size. This is simply explained by the higher redundancy of the knowledge of similar workers when their number is large: giving up a similar worker to introduce a diverse colleague entails a lower cost in terms of useful information in a large group than in a smaller one. This, together with the contribution of the diverse worker being equivalent in both cases, produces the main result that we test in the empirical part of the paper.

The predictions derived from this theoretical model, or simple extensions of it, should hold whenever producers organized in a group are trying to solve complex or intellectual tasks for which there is an objectively correct or optimal solution, as long as we can believe that workers have some capacity to discern what constitutes a good solution to the problems

they face and what does not. We test the main predictions from the theoretical model in two different contexts that satisfy these conditions, finding evidence that strongly supports the validity and quantitative relevance of the suggested mechanism, both in the lab and in the real world.

We run a lab experiment in Guinea-Bissau, with female and male Nursing and Economics university students solving a multiple-choice test twice. First each participant solves the test in isolation and then as part of a group. By exogenously allocating participants into groups of different size and composition we find causal evidence strongly supporting our main hypothesis: nurses with a single colleague do not perform better in diverse than in homogeneous groups. However, the gains from communication for nurses placed in a large group with two other nurses and an economist are about double the average gain for nurses in large homogeneous groups. Interestingly, a third treatment shows that the effect of gender diversity is larger in smaller groups, an interesting result apparently driven by discrimination towards female participants. This discrimination is a source of inefficiency and is difficult to explain assuming participants' rationality, given that male participants are likely to be paid according to the performance of that female colleague they are discriminating.

We have also shown stylized facts coherent with the main predictions from the theoretical model, using a novel administrative dataset of tax collectors in Pakistan. Tax collectors must fulfill a large number of complex tasks. This implies that those workers posted in the same district as the tax collector (Assistant Commissioner), with an important potential for knowledge spillovers, satisfy the conditions of the proposed mechanism for the definition of a group. The existence of a majority of civilian and a minority of ex-military colleagues with the same professional conditions offers a very convenient characterization of two different types of workers. Performance is measured as the amount of taxes collected in the municipality for which each tax collector is responsible, over a centrally established annual target. Tax collection is a core duty of Assistant Commissioners, giving us a comparable measure of the degree to which each AC is doing a good job. The two main findings from the study of the relationship between ACs' performance, group size and diversity support our main predictions: the performance of civilian tax collectors is increasing in the number of civilian colleagues at a diminishing rate, and working with military colleagues is associated with higher performance the larger the group in which these are placed.

Whenever workers trying to solve a complex task have the possibility of communicating with colleagues with disjoint and relevant information, and incentives are such that co-operation pays off, the proposed mechanism should be at play. Clear examples of such organizations include universities and research centers, political, bureaucratic and administrative groups, and virtually every complex enough firm in the private sector.

Many micro studies have explored whether certain kind of diversity are good or bad for performance in a broad variety of contexts, but the effect of group size has been overlooked. The main contribution of this paper is to show that the same kind of diversity might have a very different impact depending on group size, changing not only the dimension but the sign of the effect.

6 Appendix

6.1 List of assumptions for the theoretical model

This part of the appendix lists the main assumptions on which the model rests, either explicitly or implicitly explained in its description, see section 2.

- A.1.1. The cost of communication (both c and c_o) is interpreted as difficulties or time spent on communication, decreasing performance.
- A.1.2. While in most of the literature the communication costs are aggregated in the worker being asked, the receiver (see Garicano, 2000; Radner, 1993; Bolton and Dewatripont, 1994), in our paper the cost of communication falls on both workers, as communication is always bilateral.
- A.1.3. The between-type cost of communication c_o is symmetric, the same for both type-1 and type-2 workers.
- A.1.4. The between-type cost of communication is higher than the within-type costs, $c_o > 0$.
- A.1.5. Between-type costs of communication c_o and the number of workers of each type k and m in a given group, are low enough as for every agent to find it optimal to communicate with everyone in her group.
- A.2. Group composition is orthogonal to workers' effort.
- A.3. Conditional on requiring the same level of effort, every agent prefers to produce more than less.
- A.4. The knowledge of type-1 and type-2 workers is disjoint. Only type-1 workers can solve task X and only type-2 workers can solve task Y . Relaxing this assumption is equivalent to diverse workers being "less-diverse", with pretty intuitive implications.

6.2 The highest order statistic.

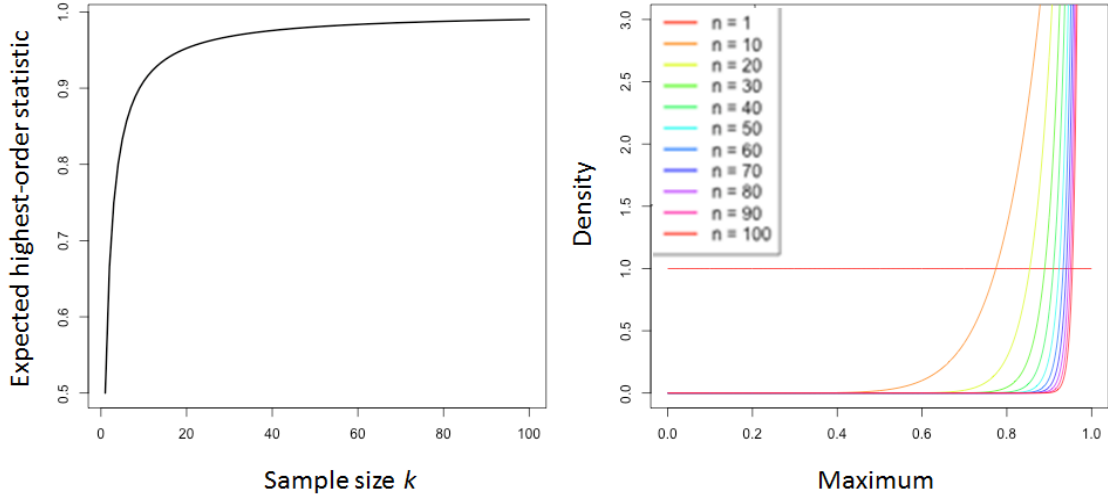


Figure 7 – Statistics from a uniform $[0,1]$ distribution. LHS: The expected or average highest order statistic for a sample of size k . RHS: The probability density function of the expected highest order statistic.

6.3 The recruitment of participants for the experiment.

The first time we asked for students availability during their exams, on sheets of paper they would sign and complete with their available dates and phone numbers. 503 out of 1104 registered students confirmed their availability to participate in the experiment during some of the four weeks of August 2018. The week with most people available resulted to be from the 30th of July to the 3rd of August, with 279 students. We decided to run the experiment on that week, with a small pilot taking place on Monday 30th of July and fixing the main experiment on Friday 3rd of August. 279 students had registered as available for that week: 67 male nurses, 53 male economists, 134 female nurses and 25 female economists. Uncertain about the proportion of absences we decided to invite a significantly larger number of students than needed (320 and 216, respectively, or 376 and 252 including the pilot) and we found that we were short of participants in each of the categories. Figure below shows the number of participants of each type required and invited for each session.⁴¹

⁴¹For the full amount of invited participants we still have to add those invited to the pilot, identical to sessions 1 and 4.

Session # (N)	Time	Allocated T		Male ECO	Female ECO	Female NUR	Male NUR	Total
Session 1 (36)	08:00	02:30	needed	4	4	14	14	36
	10:30		called	7	7	21	21	56
Session (2x36)	11:00	02:30	needed	8	8	28	28	72
	13:30		called	12	12	40	40	104
Session 3 (2x36)	14:30	02:30	needed	8	8	28	28	72
	17:00		called	12	12	40	40	104
Sessão 4 (36)	17:30	02:30	needed	4	4	14	14	36
	20:00		called	7	7	21	21	56
Total			needed	24	24	84	84	216
			called	38	38	122	122	320

Figure 8 – Table describing the number of students of each type needed and - approximately-invited to each session.

We created the list of students to be invited as follows:

1. We completed the necessary number of students in each category with randomly chosen students who have declared to be available in some other week, but not on the week of interest, after calling them to confirm if they would be willing to participate on the agreed date.
2. We managed to get the necessary number of students for each category except for female economists. We invited a few of these to participate in two sessions.⁴²
3. Once we got our sample of potential participants we randomly allocated them to the different sessions of the experiment.

After this, we wrote the lists of invited participants, and took it to their next exam at the university. This time they would confirm their availability and willingness to participate on the assigned session. We gave the other students in the same rooms the possibility of registering themselves as potential participants, as a back-up option in case we did not get enough confirmations among the invited students. This was indeed the case. Giving priority to those invited in the first place and completing vacant positions with randomly chosen students who registered their availability on this later date, we created the full list of students to be invited to each session. We still had to invite a few female economists to participate in two of them.

On the day of the pilot 41 out of 56 invited participants arrived. We were missing two female nurses to complete the "basic session" and we invited three first year nursing students who were leaving an exam to join the pilot.

On the day of the experiment, students arriving to the session to which they had been invited received a white tag when they registered for the experiment and had priority to take place in that session. Those who came to a session to which they had not been invited received a yellow tag and were kept as substitutes in case they were needed. Finally, during the morning sessions we had relatively few absences as the weather was good. As

⁴²Naturally we take care of this in the analysis. Importantly, not a single test question was included in two different sessions.

it started to rain in the afternoon and we expected a higher proportion of absences, we offered participants from previous sessions to stay and participate in later ones if necessary.

Before each session Nursing and Economics students were each directed to a different room. Once we got the number of participants of each type for the session, giving priority of the students with the white label who had not participated in any other session, the chosen participants were randomly allocated on the spot, to their corresponding room and group.

For a total of 298 observations-including 44 for the pilot-, 235 participants took part in only one session, 30 took part in two sessions and one student took part in three different sessions.

6.4 Test on the equality of medians for female and male groups

Groups	Fisher exact test (p-value)	N
H4M - H4F	0.082	48
H2M - H2F	0.701	26
D4M - D4F	0.505	36
D2M - D2F	1	16
G4M - G4F	1	36
G2M - G2F	1	21

Table 10 – This table reports the results from the non-parametric Fischer exact test for the equality of medians for *gain*, for those nurses in the female and male versions of groups with the same size and the same diversity: homogeneous (H), degree diverse (D) or degree plus gender diverse (G).

6.5 The questions for the experiment

We run five sessions of the experiment -including the pilot-. Each question has a test for economists and test for nurses. Both test have the 20 first questions in common, with the first 10 being general-knowledge questions, questions 11 to 15 corresponding to the other discipline (Economics for nurses and Nursing for economists), and questions 16 to 20 corresponding to the own discipline (Economics for economists and Nursing for nurses). The last 5 questions in each test, Nursing questions for nurses and Economics questions for economists are not included in the other type's test. Not wanting to repeat any question across different sessions, we designed 150 questions ($20 \times 5 + 10 \times 5$): 50 general-knowledge questions were designed by several of our colleagues (economists). 50 nursing questions were designed by two professors from the nursing department of the ULG and 50 economics questions were designed by a professor from the economics department at the same university. All questions were then randomized into the different sections and tests.

Examples for each group of questions include:

1. general-knowledge:
 - How many countries are there in Africa?
 - What is a water molecule composed of?
 - Which animal is impossible to find in Guinea-Bissau?

2. Economics:

- The amount of currency in circulation in the economy is determined by...
- The interest rate without discounting inflation is...
- The stock of equipment and structures used to produce goods and services is called...

3. Nursing:

- The following are factors that can produce pressure ulcers...
- The following is a basic measure for the control of contagious diseases...
- We consider "universal receiver" someone with the following blood type...

Each question had 5 possible answers, of which only one was right. Every student participating in a given session took the same test twice, first individually and then in groups. As every question was only included in one session, no student had seen or heard about a single question from the tests before participating in the experiment.

6.6 Checking for balance

Out of 21 comparisons, only the difference in the average number of correct questions for nurses in small and large degree diverse groups is statistically significant at the 10%, see table 11. This single divergence should be due to chance. A non-parametric Fischer exact test for the equality of medians does not find any statistical difference between the two groups, with a p-value of 0.76. Similarly, the average number of correct questions for economists in small degree diverse groups is 12.5, compared to 11.5 for economists in large degree diverse ones.⁴³ Hence, for degree diverse groups, the difference in the average number of correct answers in the individual test is similar and goes in the same direction for economists and nurses, implying that these small divergences should not significantly impact outcomes.

Differences across means for the nurses assigned to these groups are not statistically significant for a variety of observable demographics obtained from the survey, such as the enrollment year, age, the proportion of participants who moved to the capital for their university studies, the proportion who studied at least three years in a private school, the number of people living in the same house as the participant, or the proportion of them who have ever traveled out of the country.

⁴³A t-test for the equality of means does not find this difference significant, with a p-value of 0.37.

	Homogeneous		Degree diverse		Gender + degree	
	Group 2 (1)	Group 4 (2)	Group 2 (3)	Group 4 (4)	Group 2 (5)	Group 4 (6)
Correct, individual test	12.10 [3.05]	-.29 (.47) <i>0.532</i>	12.75 [2.70]	-1.66* (.85) <i>0.0587</i>	12 [3.53]	.06 (.89) <i>0.950</i>
Year	2.56 [.68]	-.09 (.15) <i>0.562</i>	2.81 [.75]	-.06 (.21) <i>0.775</i>	2.57 [.67]	-.02 (.17) <i>0.930</i>
Age in years	25.35 [2.77]	.43 (.78) <i>.589</i>	27.54 [1.61]	-1.19 (1.70) <i>.494</i>	24.49 [2.50]	1.30 (.79) <i>0.112</i>
Moved to capital	.52 [.51]	.06 (.12) <i>.616</i>	.375 [.5]	.097 (.15) <i>.524</i>	.55 [.51]	.16 (.14) <i>0.269</i>
Private school	.62 [.49]	-.19 (.12) <i>0.12</i>	.53 [.13]	-.13 (.16) <i>.524</i>	.52 [.51]	-.04 (.07) <i>.786</i>
People at home	10.33 [5.53]	1.39 (1.47) <i>.350</i>	10.06 [5.21]	.42 (1.58) <i>.791</i>	9.957 [6.63]	-.14 (1.76) <i>.935</i>
Traveled abroad	.43 [.50]	-.163 (.11) <i>.154</i>	.375 [.5]	-.42 (.14) <i>.791</i>	.29 [.46]	.10 (.13) <i>.433</i>
Participants	30	48	16	36	21	36

[STD], (SE), *p-values*, * $p < 0.10$

Table 11 – This table reports average characteristics for nurses in groups of size 2 and size 4, for homogeneous, degree diverse and degree plus gender diverse groups. Columns (1), (3) and (5) show average characteristics of the group of individuals specified by the column heading. Columns (2), (4) and (6) report the difference between the average characteristic of the group in the column to the left and that in the column heading.

6.7 Group size and the relative contribution of an economist for different sub-sets of questions

For each sub-set of questions the relative contribution of an economist in a large vs. in a small group:

$$\Delta(4) - \Delta(2) = [\overline{P}(\text{red squares}) - \overline{P}(\text{blue squares})] - [\overline{P}(\text{red square}) - \overline{P}(\text{blue square})] =$$

Numbers in red indicate that average performance is higher in homogeneous groups, numbers in blue that this is higher in diverse groups.

- For general-knowledge questions: $(1.78 - 0.89) - (0.63 - 1.13) = 0.89 - (-0.5) = 1.391 > 0$
- For discipline specific questions: $(2.66 - 1.35) - (1.81 - 1.26) = 1.31 - 0.55 = 0.76 > 0$
- For nursing (own-discipline) questions: $(1.1 - 1.33) - (0.43 - 1.1) = -0.23 - (-0.67) = 0.443 > 0$
- For economics (other-discipline) questions: $(1.56 - .02) - (1.375 - 0.16) = 1.54 - 1.22 = 0.32 > 0$

6.8 Distribution of correct answers for nurses in gender plus degree diverse groups

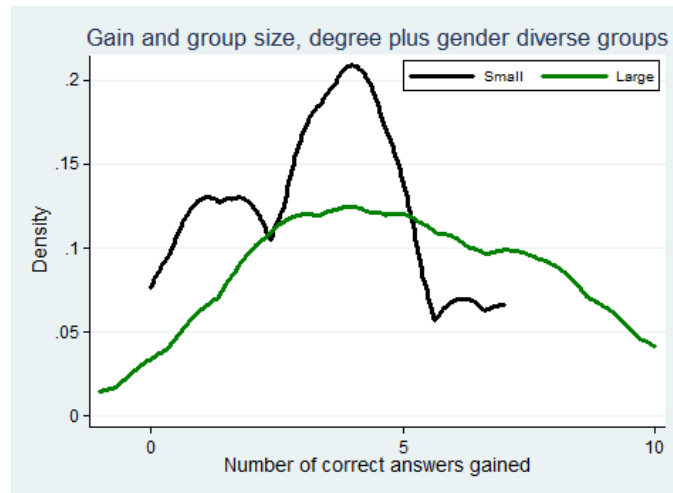


Figure 9 – Distribution of gain for nurses in gender plus degree diverse groups.

6.9 The main functions of Assistant Commissioners

Pre-1997, the functions carried out by the office of the AC fell under two main heads: Revenue and maintenance of law and order. Post Legal Reforms Act 1997, Local Government Act 2001, and various executive orders, the law and order aspect of the job has mostly been withdrawn and replaced with other miscellaneous responsibilities. The set of miscellaneous activities depends on the prerogative of the government and can change from time to time, however, they are generally allocated to all ACs in Punjab together. There can be spurts of responsibilities like an anti-encroachment drive, organizing activities around Islamic festivals or more general coordination and monitoring activities. Since AC is the linchpin of the government in the tehsil the usual practice is to refer to the AC for any queries. While not exhaustive, the following list contains many of the functions typically assigned to Assistant Commissioners in Punjab:

Official assignments:

- Revenue administration.
- Price control magistracy.
- Coordination with different state departments.
- Monitoring of devolved & non-devolved departments.

Revenue regarding functions:

- Anti-encroachment functions.
- Disposal of judicial cases.
- Monitoring of the progress of revenue officers.
- Periodically checking the revenue record maintained by the patwaris.⁴⁴
- Hearing of service appeals.
- Recruitment of lower revenue field staff.
- Revenue reporting to the higher authorities.

Coordination functions:

- Coordination with Food Department regarding procurement of wheat.
- Coordination with the Agricultural Department regarding monitoring of fertilizers and water courses.
- Inspection of development work carried out by the local government departments.
- Coordination with the Health department regarding inspections and providing support during official campaigns.
- Coordination with the local government regarding conduction of inquiries of out-dated death and birth entries.
- Coordination with the Education department regarding inspection of schools and construction of new school buildings.
- Coordination with the Irrigation department during floods.

6.10 The Punjab Agricultural Income Tax 1997

Section 4 of AIT Act, 1997 states that every person:

1. whose total agricultural income or the total agricultural income of any other person in respect of which he is assessable under this Act, for any income year (hereinafter referred to as the said income year) exceeds the maximum amount⁴⁵ which is not chargeable to tax under this Act; *or* (emphasis added)
2. who himself or any other person on whose behalf he is assessable under this Act, has, during the said income year, cultivated land measuring: (i) fifty acres or more of irrigated land; or (ii) one hundred acres or more of unirrigated land; or (iii) irrigated and unirrigated land the aggregate area of which is equal to or more than fifty acres of irrigated land, one acre of irrigated land being reckoned as equivalent to two acres of unirrigated land, shall file a return of his total agricultural income or the agricultural income of such other person, as the case may be, for the said income year in such form and by such date as may be prescribed.

⁴⁴A patwari or *rural accountant* is the lowest state functionary in the Revenue Collection system. His or her job encompasses visiting agricultural lands and maintaining record of ownership and tilling, which has historically given them a very large influence in local communities. They have their primary base in rural areas.

⁴⁵The Punjab Agricultural Income Tax Rules 2001 states that this threshold is Rs. 80,000.

6.11 The independence of tax targets from previous tax collection

	Tax Targets (Rs. in million)	
	(1)	(2)
Past district tax collected	0.01 (0.10)	0.05 (0.21)
Election year	-2.77 (3.65)	0.40 (5.37)
Real wage	-0.01 (0.05)	-0.01 (0.06)
Population estimates	-0.00 (0.00)	-0.00 (0.00)
Rural employment	0.32 (0.21)	0.28 (0.30)
Agriculture production	0.00*** (0.00)	0.00*** (0.00)
Irrigated area	0.02*** (0.01)	0.04** (0.02)
Year FE	Yes	Yes
District FE	No	Yes
Observations	120	120

Standard errors in parenthesis * p<0.10 ** p<0.05 *** p<0.01

Table 12 – Determinants of tax targets. This table shows the determinants for annual targets at the district level. As we can see, only irrigated area and agriculture production (measured as land area owned) are statistically significant. Taxes collected in the district in the previous year are not.

6.11.1 The original documents

The personal and professional information about each civil servant has been extracted from the ‘Individual Career Planning Chart’, the file used by the PAS to keep track of their employees. The first part of this document contains a picture of the worker, its id. number, name and date of birth, and a complete list of the postings and stations the worker has been allocated to along its career in the civil services, together with her supervisors’ evaluation of her performance in the post and her fitness for promotion, see figure 10 in appendix 6.11.2. The second part contains personal information, such as the domicile, the district of origin and date of birth, the religion, the academic and professional titles, training received, languages spoken, marital status, the occupational group service, or the date of promotion to each of the different grades, see figure 11.

In order to complete the dataset, the data extracted from this document has been matched with the measure of performance, only available for ACs. This measure, given by the amount of agricultural tax collected in a tehsil in the corresponding month divided by the annual tax target, as described below, was extracted for each case from one of two

documents. The first of them is the ‘Statement showing the recovery position of agricultural income’ for a given tehsil and month, see figure 12 in appendix 6.11.3. It contains the target tax —based on the official record of size and number of farmlands —for the financial year and the recovery during the month, among other measures related to the target tax and the amount of taxes collected. This form is completed by the corresponding AC in charge of that tehsil at that given period, and signed by her supervisor (DC). The second document consists of a listing of the tax collected in the corresponding month in each of the tehsils in a given district, see figure 13 in appendix 6.11.3. This document is completed by the Deputy Commissioner (DC) using the statement handed to her by each of the ACs she supervises. As can be seen there are two kinds of targets: one is for the current financial year and the other is for arrears from all the past financial years. For the purpose of this study we restrict attention to the performance against targets set for the current financial year.⁴⁶ Tax collected each month is under the title ‘current recovery’ or under ‘recovery during the month’.

⁴⁶Since the current financial year is the main priority of the government, ACs expend more effort trying to meet these targets. Annual targets are provided in the column titled ‘demand’.

6.11.2 Officers' career charts



No. _____

Name _____

Date of birth _____

*Post _____

*Date of present appointment _____

Field

Secretariat

Corporation

Soft

Hard

Foreign

SERVICE

Year	Year of Service	Post held	Class/Grade	Department /	Station	ACR Assessment		Fitness for Promotion
						Reporting Officer	Counter-signing Officer	
1975	1st	E.A.C. 10-1-75 to 31-12-75	17	S & G.A.D.	KASUR	AVG.	GOOD	PREMATURE
1976	2nd	" 31-12-75 to 1-1-76	"	"	"	AVG.	AVG.	PREMATURE
"	"	" 1-1-76 to 31-12-76	"	"	"	GOOD	GOOD	PREMATURE
"	"	" 31-12-76 to 1-1-77	"	"	"	GOOD	V.GOOD	FIT.
1977	3rd	" 1-1-77 to 14-4-77	"	"	"	GOOD	GOOD	"
"	"	" 14-4-77 to 31-12-77	"	"	"	GOOD	GOOD	FIT.
"	"	" 31-12-77 to 1-1-78	"	"	"	GOOD	GOOD	FIT.
1978	4th	" 1-1-78 to 4-2-78	"	"	"	GOOD	GOOD	FIT.
"	"	" 4-2-78 to 31-12-78	"	"	LAHORE	GOOD	AVG.	IN DUE COURSE
"	"	" 31-12-78 to 1-1-79	"	"	"	V.GOOD	GOOD	FIT.
1979	5th	" 1-1-79 to 1-1-80	"	"	"	V.GOOD	GOOD	FIT.
1980	6th	" 1-1-80 to 1-1-81	"	"	"	AVG.	AVG.	NOT YET
1981	7th	" 1-1-81 to 31-12-81	"	"	WAZIRABAD	V.GOOD	GOOD	FIT.
"	"	" 31-12-81 to 1-1-82	"	"	"	V.GOOD	GOOD	FIT.
1982	8th	" 1-1-82 to 31-12-82	"	"	"	V.GOOD	GOOD	FIT.
1983	9th	" 31-12-82 to 1-1-83	"	"	"	V.GOOD	AVG.	FIT.
1984	10th	" 1-1-83 to 1-1-84	"	"	"	V.GOOD	GOOD	FIT.
"	"	" 1-1-84 to 31-12-84	"	"	HANIKMAN	V.GOOD	GOOD	FIT.
1985	11th	" 31-12-84 to 1-1-85	"	"	"	V.GOOD	GOOD	FIT.
"	"	" 1-1-85 to 31-12-85	"	"	CHINHOT	GOOD	GOOD	FIT.
"	"	" 31-12-85 to 1-1-86	"	"	"	GOOD	V.GOOD	FIT.
1986	12th	" 1-1-86 to 31-12-86	"	"	"	HE DID NOT WORK UNDER ANY R.O.		
1987	13th	" 31-12-86 to 1-1-87	"	"	"	PERIOD LESS NO REPORT		
"	"	" 1-1-87 to 31-12-87	"	"	CHINHOT	PERIOD LESS NO REPORT		
"	"	" 31-12-87 to 1-1-88	"	"	KASUR	GOOD	GOOD	FIT.
1988	14th	" 1-1-88 to 31-12-88	"	"	"	Not in CR		
"	"	" 31-12-88 to 1-1-89	"	"	"	NO REPORT		
"	"	" 1-1-89 to 31-12-89	"	"	EDUCATION LAHORE	PERIOD LESS NO REPORT		
1989	15th	" 31-12-89 to 1-1-90	"	"	"	"		

*Entries with lead-pencil.

Figure 10 – Official civil servant career chart, sheet with the postings over his professional career.

CONFIDENTIAL 4-7
8/4/7

INDIVIDUAL CAREER PLANNING CHART

1. No. _____

2. Occupational Group Service **EX. PCS (P.B.)**
DIRECT RECRUIT (L.M.)

3. Name _____

4. Seniority Position **145**

5. Date of joining Govt. service _____

6. Religion **ISLAM**

7. Date of birth _____

8. Appointment with grade **E.A.C. (17)**

9. Marital status **MARRIED**

10. Date of superannuation _____

11. Domicile **PUNJAB**

12. Service particulars of spouse

13. Home District **BAHAWALPUR**

14. Medical category _____

15. Qualifications

Academic	Professional	Languages known
B.A. (POLITICAL SCIENCE)	LL.B.	ENGLISH, URDU, & PUNJABI

16. Training received

	Name of institution attended	Country	Duration		Particulars of the course
			From	To	
Domestic					REVENUE
Foreign					

17. Countries visited

Country	Duration		Purpose
	From	To	

18. Merit position in

	Year	Position
P.P.S.C Competitive Examination	_____	_____
P.P.O Examination	_____	_____
Any other Examination	_____	_____

19. Date of entry/promotion

Grade	16	17	18	19	20	21	22	23
Temporary								
Substantive		6.1.1975	16.4.88					

*Entries with lead-pencil

Figure 11 – Official civil servant career chart, sheet with his personal, education and professional information.

6.11.3 The tax collection form

(نہایت محترم)

STATEMENT SHOWING THE RECOVERY POSITION OF AGRICULTURAL INCOME TAX
UNDER HEAD 011630001173 FOR THE MONTH OF December, 2007 District D.G.KI

Head of Account No. 011630001173	Demand	Remission	Suspension	Net Demand	Previous Recovery	Recovery during month	Total recovery	Balance
A.I.T. (Previous)	9664766	—	6368392	3296374	2482954	114322	2597276	699098
A.I.T. (Current)	—	—	—	—	—	—	—	—
Total	9664766	—	6368392	3296374	2482954	114322	2597276	699098

verified for Rs. 114322/- (One lac, fourteen thousand = three hundred & twenty two only)

District Officer (Revenue)
Dera Ghazi Khan

Figure 12 – Official form for the communication of the tax collected in a given month.

AGRICULTURAL INCOME TAX, DISTRICT MUZAFFARGARH,
FOR THE MONTH OF September, 2007.

PREVIOUS A-I-T.

S. No.	Name of Tehsil	Demand	Suspension	Net Demand Recoverable	Previous Recovery	Current Recovery	Total Recovery	Balance	Percentage Month	Total
1-	M. Garh	17102682	—	17102682	76650	9300	85950	17016732	—	1%
2-	Kot Addu	28353571	—	28353571	87793	38100	125893	28227678	—	—
3-	Alipur	2079273	—	2079273	34150	44706	78856	2000417	2%	4%
4-	Talari	18396542	—	18396542	50010	9500	59510	18337032	—	—
Total A		65932068	—	65932068	248603	101606	350209	65581859	—	1%

CURRENT A-I-T.

1-	M. Garh	—	—	—	—	—	—	—	—	—
2-	Kot Addu	—	—	—	—	—	—	—	—	—
3-	Alipur	—	—	—	—	—	—	—	—	—
4-	Talari	—	—	—	—	—	—	—	—	—
Total B		—	—	—	—	—	—	—	—	—
G.Total A+B		—	—	—	—	—	—	—	—	—

Figure 13 – Official form for the registry of agricultural income tax.

6.12 Average performance per month

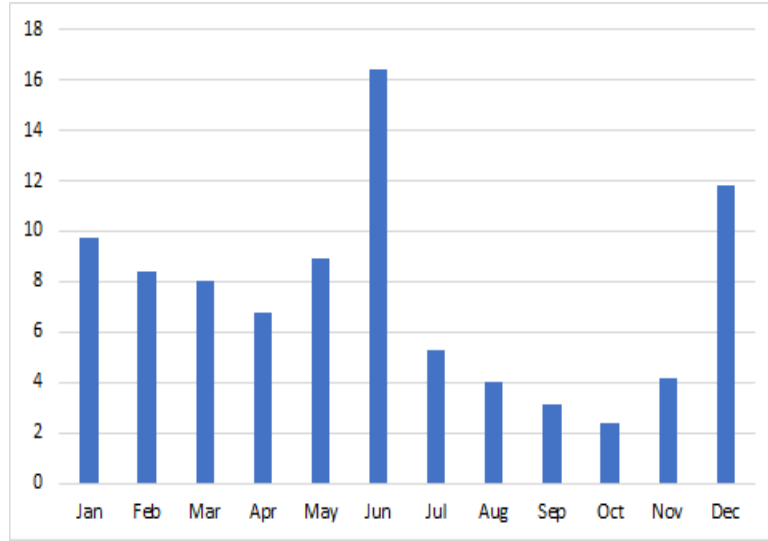


Figure 14 – Average performance in the different months of the year.

6.13 Marginal effect of an extra colleague on individual performance

With the coefficient associated to the variable *colleagues* $\beta_1 = 0.241$ and the coefficient associated to the variable *colleagues*⁴ $\beta_2b = -1.06 \times 10^{-7}$ in column (2), table 8, the marginal effect of having an extra colleague on *performance* is given by

$$\frac{\partial performance}{\partial colleagues} = 0.241 - 4 \times 1.06 \times 10^{-7} colleagues^3.$$

The following table shows this marginal effect for groups of different size, between 1 and 80.⁴⁷

<i>colleagues</i>	1	5	15	25	50	75	80
$\frac{\partial Y_{ikdmt}}{\partial colleagues}$	0.241	0.2409	0.239	0.234	0.188	0.062	0.024

Table 13 – Marginal effect of an extra colleague on individual performance, in homogeneous groups.

⁴⁷The 97.2% of all observations for homogeneous groups are in groups with 80 or less workers. In concrete, out of 5428 observations, 5376 correspond to civilian tax collectors in groups with a size lower than 80.

6.14 The number of military colleagues for the civilian ACs in our sample

Number of ex-military colleagues	Freq.	Percent	Cum.
0	5,427	78.26	78.26
1	1,213	17.49	95.75
2	201	2.90	98.64
3	82	1.18	99.83
4	9	0.13	99.96
5	3	0.04	100.00
Total	6,935	100.00	

Figure 15 – Distribution of the variable *military colleagues* for the civilian ACs in our sample.

References

- Akerlof, George A and Rachel E Kranton (2005), “Identity and the economics of organizations.” *The Journal of Economic Perspectives*, 19, 9–32.
- Alesina, Alberto and Eliana La Ferrara (2000), “Participation in heterogeneous communities.” *The quarterly journal of economics*, 115, 847–904.
- Alesina, Alberto and Eliana La Ferrara (2005), “Ethnic diversity and economic performance.” *Journal of economic literature*, 43, 762–800.
- Alvarez Pereira, B, TJ Correia, FY Kumah, and B Na salù (2017), “Guinea-bissau: Selected issues paper, chapter 4: Impact of economic structure and policy on poverty incidence and distribution.” *IMF Country Report No. 17/381, November*.
- Angrist, Joshua D (2014), “The perils of peer effects.” *Labour Economics*, 30, 98–108.
- Antràs, Pol, Luis Garicano, and Esteban Rossi-Hansberg (2006a), “Offshoring in a knowledge economy.” *The Quarterly Journal of Economics*, 121, 31–77.
- Antràs, Pol, Luis Garicano, and Esteban Rossi-Hansberg (2006b), “Organizing offshoring: Middle managers and communication costs.”
- Bachman, Jerald G, Lee Sigelman, and Greg Diamond (1987), “Self-selection, socialization, and distinctive military values: Attitudes of high school seniors.” *Armed Forces & Society*, 13, 169–187.
- Bertrand, Marianne, Robin Burgess, Arunish Chawla, and Guo Xu (2016), “The costs of bureaucratic rigidity: Evidence from the indian administrative service.”
- Bertrand, Marianne and Esther Duflo (2017), “Field experiments on discriminationa.” In *Handbook of Economic Field Experiments*, volume 1, 309–393, Elsevier.
- Bolton, Patrick and Mathias Dewatripont (1994), “The firm as a communication network.” *The Quarterly Journal of Economics*, 109, 809–839.
- De La Cal, J and J Cárcamo (2005), “Inequalities for expected extreme order statistics.” *Statistics & probability letters*, 73, 219–231.

- Deininger, Klaus et al. (2001), "Land policy and administration: Lessons learned and new challenges for the bank's development agenda." *World Bank*.
- Enemark, Stig and Paul Van der Molen (2008), *Capacity assessment in land administration*. FIG Denmark.
- Gagliarducci, Stefano and M Daniele Paserman (2011), "Gender interactions within hierarchies: evidence from the political arena." *The Review of Economic Studies*, rdr046.
- Garicano, Luis (2000), "Hierarchies and the organization of knowledge in production." *Journal of Political Economy*, 108, 874–904.
- Haltiwanger, John C, Julia I Lane, and James Spletzer (1999), "Productivity differences across employers: The roles of employer size, age, and human capital." *American Economic Review*, 89, 94–98.
- Hansen, Zeynep, Hideo Owan, and Jie Pan (2006), "The impact of group diversity on performance and knowledge spillover—an experiment in a college classroom."
- Hoxby, Caroline (2000), "Peer effects in the classroom: Learning from gender and race variation."
- Jackson, Joshua J, Felix Thoemmes, Kathrin Jonkmann, Oliver Lüdtke, and Ulrich Trautwein (2012), "Military training and personality trait development: Does the military make the man, or does the man make the military?" *Psychological Science*, 23, 270–277.
- Lazear, Edward P (1998), "Diversity and immigration." Technical report, National Bureau of Economic Research.
- Lazear, Edward P (1999a), "Culture and language." *Journal of Political Economy*, 107, S95–S126.
- Lazear, Edward P (1999b), "Globalisation and the market for team-mates." *The Economic Journal*, 109, 15–40.
- Manski, Charles F (1993), "Identification of endogenous social effects: The reflection problem." *The review of economic studies*, 60, 531–542.
- Nasim, Anjum (2012), "Agricultural income taxation: estimation of the revenue potential in Punjab." *The Pakistan Development Review*, 321–335.
- Ozer, Daniel J and Veronica Benet-Martinez (2006), "Personality and the prediction of consequential outcomes." *Annu. Rev. Psychol.*, 57, 401–421.
- Pelled, Lisa Hope, Kathleen M Eisenhardt, and Katherine R Xin (1999), "Exploring the black box: An analysis of work group diversity, conflict and performance." *Administrative science quarterly*, 44, 1–28.
- Qazi, Mohammad Usman (2005), "Social Assessment of Land Record Management Information System Programme." *Background Paper, The World Bank Pakistan Country Office, Islamabad, Pakistan*.
- Qazi, Mohammad Usman (2006), "Computerization of land records in Pakistan." *Islamabad, LEAD International*.
- Radner, Roy (1993), "The organization of decentralized information processing." *Econometrica: Journal of the Econometric Society*, 1109–1146.

- Rodan, Simon and Charles Galunic (2004), "More than network structure: How knowledge heterogeneity influences managerial performance and innovativeness." *Strategic management journal*, 25, 541–562.
- Steiner, Ivan D (2007), "Group process and productivity (social psychological monograph)."
- Williams, Katherine Y and Charles A O'Reilly III (1998), "A review of 40 years of research." *Res Organ Behav*, 20, 77–140.