Diversity and Team Performance in a Kenyan Organization*

Benjamin Marx
Vincent Pons
Tavneet Suri

June 2015

Abstract

We present findings from a field experiment on team diversity. Individuals working as canvassers in an elections-related experiment were randomly assigned a junior teammate, a senior manager and a set of households to visit. This created random variation in the degree of horizontal diversity, vertical (or hierarchical) diversity and external (or client) diversity within each team. We find that ethnic diversity among teammates decreases team performance, while diversity along the vertical dimension improves performance. The data on time use suggests that horizontally homogeneous teams organized tasks in a more efficient way, while vertically homogeneous teams exerted lower effort.

Keywords: ethnic diversity, organizational behavior, labor management, performance

JEL Classification: D22, J24, L22, M54, O12

*We are extremely grateful to Suleiman Asman, Eleanor Wiseman, Geoffrey Onyambu and Bonnyface Mwangi for exceptional research management and assistance in the field. We benefited from helpful comments and suggestions from Anna Aevarsdottir, Evan Apfelbaum, Abhijit Banerjee, Nick Bloom, Ray Fisman, Christina Patterson, Jonathan Petkun, Otis Reid, Michael Stepner, as well as seminar audiences at the NBER Summer Institute, Georgetown, MIT, MIT Sloan, Novafrica, the Rotman School of Management, University of Capetown, UVA Batten School, University of Bocconi and IIES Stockholm. We gratefully acknowledge financial support from MIT Sloan School of Management and the Program on Innovation in Markets and Organizations at MIT Sloan. Marx is at the Department of Economics, MIT, Pons is at Harvard Business School and Suri is at the MIT Sloan School of Management. Suri is the corresponding author: E62-517, 100 Main Street, Cambridge MA 02142. Email: tavneet@mit.edu.

1 Introduction

A central question in organizations is whether there exists an optimal balance between diversity and homogeneity within teams of workers. While diversity brings in a wider range of skills and ideas, it also creates communication costs and other frictions (via perhaps trust issues, lower job satisfaction) inside the organization – many studies have highlighted this trade-off (Lazear (1999), Prat (2002), Hamilton et al. (2003), among others). What is the level of diversity that maximizes team performance? Does team diversity make management more effective, or facilitate relationships with clients?

In this paper, we study these questions in the context of a field experiment conducted inside a Kenyan organization. A key contribution of our design is that we are able to study, within a single organization, the effect of three different dimensions of diversity. First, we analyze the impact of horizontal diversity (diversity among individuals holding the same job or position) on various measures of team performance. Second, we look at the effect of hierarchical or "vertical" diversity (between managers and more junior workers) on performance. Third, we test whether similarity between team members and the clients they interact with affects performance. We present results pertaining to three types of diversity (by ethnicity, gender and age), and we emphasize the effect of ethnic diversity, given these effects are much larger and more robust than the effects of other forms of diversity.[1,2]

Our field experiment was conducted inside a Kenyan non-profit research organization that collects survey data. The experiment focuses on a standard set of tasks that employees of this organization conduct, and that we use to measure performance. The organization resembles other survey companies across the world. A sizeable fraction of its workforce consists of short-term staff whose main task is to enumerate or survey households and individuals, and substantial resources are invested in the monitoring of this short-term staff. Making management (or supervision) and quality controls effective is therefore crucial to the success of organizations such as the one we study.

In 2012, in partnership with this organization and the Independent Electoral and Boundaries Commission of Kenya (IEBC), we conducted a voter canvassing exercise around the forthcoming Kenyan general election. The IEBC had a mandate to create a new register of eligible voters across the country. Our canvassing experiment aimed to increase voter registration in one disenfranchised Nairobi neighborhood, called the Kibera slum. The intervention covered approximately 15,000 households in 300 enumeration areas (census tracts)[3] and provided residents with

[1] There is a voluminous literature on the effects of ethnic diversity in Sub-Saharan Africa. Alesina and Ferrara (2005) provide an excellent review.

[2] We cannot look at measures of diversity in skills or education, as all the staff members had approximately the same level of education. Papps et al. (2011) and Hamilton et al. (2012) look at the effect of diversity in ability on team performance.

[3] This is the size of our treatment sample. There was an equal-sized control group where households did not

information on the registration process via door-to-door canvassing conducted by teams of two.

Subsumed within this registration experiment, we designed a field experiment to study how diversity within teams and along the organizational hierarchy affects effort and performance. This involved three levels of randomization. First, we randomly allocated the canvassers to teams of two. Second, we randomly allocated groups of teams (approximately four teams each) to a manager responsible for monitoring these teams in the field. This created variation in what we call vertical diversity, i.e., in the degree of heterogeneity between managers and the junior workers they supervised. Third, each team was randomly allocated a set of enumeration areas (henceforth EAs) to visit in a given random order. Fieldwork lasted for two weeks and teams covered a set of ten EAs each week. This introduced variation in the degree of similarity between canvassers and "clients", i.e. the households they were expected to visit.

We measure team performance in two ways. First, we observe whether a particular canvassing visit was completed, i.e. whether the correct household was found and whether the pair of canvassers was able to complete their task. Second, we look at the duration of each visit. Canvassers reported the time at which they started canvassing each household, and the time the visit ended. We treat longer durations as a positive outcome, indicating both effort on the part of canvassers and active involvement of the target households, as very few of the durations exceeded what we expected as the typical visit length.[4]

For both measures of performance, we find that ethnically homogeneous teams on the horizontal dimension perform better. This effect is large in magnitude – more than twice the size of the effect of homogeneity by gender or age. We find the opposite effect for vertical homogeneity. Teams with a manager whose ethnicity differs from that of the two teammates perform better, suggesting that even though horizontal homogeneity enhances performance, vertical homogeneity is counterproductive in our context, possibly due to distortions of managerial effort (as in Bandiera et al. (2009)). Finally, we find that the performance of teams is unaffected by observable similarities with the households on their canvassing list. Given the prevalence of ethnic voting in Kenya, this is a surprising finding as we had expected that the ethnicity of the households would matter for teams' effort and performance. Instead, we find that canvassers did not strategically target households of the same ethnicity, nor did they spend more time canvassing these households.

We provide evidence for the mechanisms underlying these effects. First, we construct measures of daily team effort and performance. We compute the time spent in the field each day, the actual time devoted to canvassing households, the time spent walking between dwellings (the last two add up to the first), and the number of successful visits conducted each day. We

receive any canvassing visit.

[4]Later in the paper, we document that visit duration is a monotonic function of both horizontal homogeneity (increasing) and vertical homogeneity (decreasing).

interpret time spent in the field as a measure of effort, and the time split between canvassing and non-canvassing, as well as the number of successful visits as measures of productivity. We find that ethnically homogeneous teams (on the horizontal dimension) spend the same amount of time in the field, more time canvassing, less time walking or searching for households, and hence also complete more visits. These teams are also more likely to divide up tasks – they conduct a higher fraction of visits individually. While completing the same number of visits on average, vertically homogeneous teams spend less time in the field and less time canvassing households, implying lower effort levels in those teams.

Second, we conducted a survey of the staff to better understand team dynamics. We present evidence that the working atmosphere was more cohesive within ethnically homogeneous teams. Canvassers in a homogeneous pair were less likely to complain about their teammate, and less likely to report that they performed better than their teammate. There is no evidence that horizontal homogeneity improved work ethic or (self-reported) hours. Looking at vertical homogeneity, we find that canvassers in these teams have a more negative assessment of their own performance and report working less hours, indicating a loosening of discipline and lower effort. Finally, the survey data indicates that workers neither have a (stated) preference for working with co-ethnic teammates, nor do they speak a different language with teammates or managers of the same ethnicity while working.

1.1 Contribution to Literature

There is a large literature on the role of diversity and/or homogeneity in teams and organizations, both in economics as well as in psychology and organizational behavior. Reviewing the literature outside of economics is beyond the scope of our paper, but good discussions can be found in Apfelbaum et al. (2014) and Phillips and Shim (2011). In economics, the question of whether and how diversity matters has been studied across fields, in macroeconomic as well as microeconomic contexts. In this section, we highlight some of the studies in these areas and discuss the contributions of our paper to the existing literature on diversity in firms and organizations.[5]

The early literature on diversity and performance was largely theoretical and descriptive. Becker (1957) argued that employees of a firm have preferences in terms of whom they want to

[5]Aside from the evidence we discuss on firms and organizations, there has been a lot of work on the role of ethnic diversity in other economic spheres. At the macro level, there is a literature documenting the strong negative correlations between diversity and measures of economic performance across countries, within countries, and across U.S. states (Alesina and Ferrara (2005) provide an excellent review). At the micro level, Fafchamps (2000) and Fisman et al. (2012) document the role of ethnic diversity in credit market interactions; Anwar et al. (2012) and Shayo and Zussman (2011) look at legal systems in the U.S. and in Israel, respectively; Fairlie et al. (2014) study schools in the U.S.; Price and Wolfers (2010) look at NBA refereeing. There is also a large literature showing a negative correlation between ethnic diversity and public goods provision (Easterly and Levine (1997), Alesina et al. (1999), and Miguel and Gugerty (2005)), among others).

work with – workers incur disutility from working with individuals from other groups. This "taste for discrimination" can explain why wage differentials arise between ethnic groups. Discriminators are compensated for working with the groups they do not like, and hence receive higher wages. Arrow (1973) extended this approach with a model of statistical discrimination, which assumes that employers have beliefs about the ability of different groups. Lazear (1999) and Prat (2002), among others, have discussed the conditions under which homogeneity is optimal, but the empirical evidence on these questions is still limited.

There exists a "business case for diversity" positing that diverse teams can serve a broader spectrum of customers and solve a wider range of problems (Hamilton et al. (2012)). This argument posits that ethnic diversity brings in diversity of skills – different groups often have access to different (disjoint) information and skill sets, generating complementarities across groups. Nevertheless, for such complementaries to arise, the skills possessed by different groups must be relevant to one another, and there must exist opportunities for learning (Lazear (1999)). Consequently, the magnitude of the positive relationship between diversity and productivity should increase when groups are more diverse along the skill dimension, and decrease when groups are more diverse along other dimensions, such as language (Hamilton et al. (2012)).[6] Furthermore, diversity can improve the functioning of markets by reducing herding behavior (Levine et al. (2014)).

Hjort (2014) studies the effect of ethnic diversity on team productivity in the Kenyan flower industry. He finds that diversity adversely affects productivity, and that ethnic conflict accentuates this negative effect. He argues that these findings are consistent with preferences à la Becker. His results, however, are at odds with those of Hamilton et al. (2003), and with the more recent experimental results in Hoogendoorn et al. (2012) and Hoogendoorn and van Praag (2014), who study the impact of ability dispersion and ethnic diversity on team performance. The latter study finds that a large degree of diversity within teams enhances productivity in a sample of MBA students and young entrepreneurs in the Netherlands. In a related paper, Lyons (2014) studies diversity by nationality in teams within oDesk (an online labor market) in a field experiment. She finds a negative effect of team homogeneity on performance.

Furthermore, Lang (1986) argues that preference-based or discrimination arguments are not borne out empirically. He frames the question of ethnic diversity in terms of the costs that verbal and non-verbal communication between members of different groups imposes to the firm. Zenger and Lawrence (1989) similarly show that age differences increase communication costs between employees. In a related experiment, Reinhard and Warglien (2007) investigate the conditions under which teammates are able to develop a common language. They find that a

[6]The findings of Hamilton et al. (2012) are broadly in line with these predictions: using personnel records from a Californian garment factory, they find that teams heterogeneous in ability are more productive, while more diverse teams (in terms of age and ethnicity) are less productive after controlling for ability.

functional working language (code) is more likely to arise if teammates possess *ex ante* a larger repertoire of common linguistic symbols. These technological advantages are augmented with the higher ability of homogeneous teams to prevent free-riding by imposing sanctions – Kandel and Lazear (1992) show this formally. Diversity, on the other hand, tends to weaken social ties and to reduce peer pressure (Hamilton et al. (2012)), which in turn hampers the ability of teams to prevent free-riding (Mas and Moretti (2009)).

In this paper, we contribute to this literature along a number of dimensions. First, the literature has generally focused on the impact of horizontal homogeneity between coworkers or teammates, rather than the impact of homogeneity across the firm hierarchy. This is at odds with the large literature (e.g., Lazear et al. (2014)) highlighting the key role managers play in determining team performance. One important exception is Bandiera et al. (2009), who find that similarities (social connections) between managers and workers increase individual productivity in the cross-section (under fixed wages), but also that an increase in the *average* level of connections can decrease firm productivity by inducing distortions in the allocation of managerial effort. Our contribution is therefore to study both horizontal as well as vertical diversity within the same organization.[7] In addition, we study how external homogeneity, i.e. similarities between workers and their clients, affects worker performance.

Second, the existing literature does not shed much light on the specific mechanisms through which diversity/homogeneity can become consequential. There are a number of potential mechanisms described in the literature, including preferences; communication technologies (homogeneous teams face lower communication costs); norms and behavior enforcement within groups; and the notion that diversity begets ideas and innovation. In this paper, we collected measures of not just performance, but also of effort. We conducted a short survey of the workers to better understand the mechanisms underlying the effects we find. We show that horizontally homogeneous teams (along the ethnicity dimension) select a different organization of work and ultimately deliver higher performance. We also find that these teams are more cohesive. Looking at vertical homogeneity, we find that teams where one teammate is of the same ethnicity as the manager work less hours. This suggests that managers of the same ethnicity allow teams to shirk more as they are potentially monitored less intensively.

Third, we contribute to the identification of the effects of team diversity on effort and performance by using a field experiment to generate random variation in team diversity. Because of this randomized design, we are able to identify the causal impacts of diversity along the three dimensions of diversity (horizontal, vertical, external) discussed above.

The rest of the paper is organized as follows. Section 2 describes our experiment and section

[7] Hjort (2014) uses quasi-experimental variation in both "upstream" and "downstream" ethnic interactions, but in his context the upstream/downstream dimension refers to the supply chain, rather than an actual reporting hierarchy within the firm.

3 describes our data. We present the econometric framework in section 4 and our main results in section 5. Section 6 provides a discussion of these results using data on time use and evidence from a survey of the canvassing staff. Section 7 concludes.

2 Experimental Design

2.1 Context and Experimental Site

A general election was held in Kenya in March 2013. Prior to this election, in the aftermath of a heavily contested general election in 2007, a new electoral commission, the IEBC, was set up to demarcate electoral boundaries and organize future ballots. As part of its mandate to run more transparent elections, the new commission undertook to re-register the entire Kenyan electorate using biometric equipment. The registration process took place across the entire country in one month between November and December 2012. To help the IEBC address issues of credibility given the 2007 election and accomplish its mission to register all eligible voters, we collaborated with them to implement a door-to-door canvassing experiment during the voter registration drive in 2012.[8]

We conducted this experiment in Kibera, the largest of Nairobi's slums. Kibera was chosen as the site of the experiment for two reasons. First, it is host to a large migrant population, and there is evidence that poor and migrant voters suffer the most from the various obstacles to registration (Braconnier et al. (2014)). Kibera had also been a focal point in the 2007 post-election violence. Second, building on earlier fieldwork conducted in the slum (Marx et al. (2014)), we had data available on the households residing in Kibera, including a geo-localized census of residents, the full micro data from the 2009 national census, and EA maps for the entire area. A more detailed description of how these data were collected is provided in Section 3. Figure 1 shows a map of the Kibera area with the EAs demarcated.

2.2 Ethnic Diversity in Kenya

There are over 30 different ethnic groups in Kenya, as defined in the official census, including five main groups representing over sixty percent of the population: the Kikuyus (17%), Luhyas (14%), Kalenjins (13%), Luos (10%), and Kambas (10%) (figures from the 2009 Kenyan census). These groups differ in terms of their linguistic proximity with each other – the Kikuyus, Luhyas

[8]This experiment was designed to contribute to the literature on voter education campaigns. A number of studies have investigated the role of voter information campaigns in developing countries (see Aker et al. (2011) and Banerjee et al. (2010), among others), but there is little evidence on the effect of these campaigns in contexts where electoral institutions are entirely new. In addition, while experiments on voter registration have been conducted in advanced economies (see Braconnier et al. (2014) for a study in France), we are not aware of any such study in a developing country context, even though registration might also represent a significant barrier to electoral participation in these countries.

and Kambas are Bantu speaking peoples,[9] while the Luos and Kalenjins belong to the Nilotic linguistic family. Political coalitions between these groups have also varied over time. Three of the four Kenyan presidents hailed from the majority Kikuyu tribe, but the longest-serving Kenyan president, Daniel Arap Moi, belonged to the Kalenjin tribe. The two most recent elections pitted a member of the Kikuyu tribe (Mwai Kibaki in 2007 and Uhuru Kenyatta in 2013) against a member of the Luo tribe (Raila Odinga in 2007 and 2013). However, given there are five large tribes, ethnic coalitions are more complex. For example, Kalenjins sided with the Luo (and against the Kikuyu) candidate in 2007, but then sided with the Kikuyu candidate (and against the Luo candidate) in 2013.

The ethnic composition of the site of the experiment, the Kibera slum, is not reflective of that of Kenyan society overall. Luos and Luhyas are the most prevalent groups in Kibera (together constituting 63% of the slum population, based on our 2012 census), while Kikuyus are a minority locally (6%).

2.3 Canvassing Experiment

2.3.1 EA Randomization

The voter registration experiment was designed to tease apart the effects of pure information on the registration process versus those of specific information about the new electoral commission, both of which were disseminated via door-to-door canvassing. The experimental design included two treatment groups. In the first group, households were encouraged to register and given information about the location of registration booths in their neighborhood, and what documents were required to register. In the second group (henceforth the IEBC treatment), households were, in addition, given detailed information about the IEBC, in particular about its record in organizing by-elections and its efforts to establish a reliable voter register via biometrics.

The experimental treatment was clustered at the level of the EA (EAs in Kibera typically correspond to a compact block of dwellings within the slum). There are 643 EAs in the slum, 603 of which were part of our sampling frame for the experiment. To maximize power to test the effects of any canvassing visit on registration, 303 EAs were randomly chosen to be part of the control group, after stratifying by village and by being above or below the median EA population. Half of the remaining EAs were allocated to each treatment group. Overall, the experiment covered a total of 31,646 households in 603 EAs (the full sample of households in the 2012 census we collected), with a total of 15,676 households treated in the 300 treatment EAs.

[9] Swahili, the main language spoken throughout Kenya, is itself a Bantu language.

2.3.2 Implementation

The fieldwork for this experiment was entrusted to a Kenyan organization affiliated with a U.S.-based non-profit research institution. The Kenyan organization has conducted field operations in Nairobi and the rest of the country for over ten years. The organization typically hires long-term research coordinators and research assistants, as well as short-term survey staff (though some of these staff may be re-hired to work on other projects, based on performance). The effort and performance levels of the survey staff are a major determinant of the organization's overall productivity. Monitoring the performance of the short-term staff requires high-frequency supervision and quality controls, as is typically the case in survey companies across the world.

To implement the experiment, the organization recruited sixty canvassers for a period of three weeks, including the time allocated to training. Sixteen of these canvassers had some prior fieldwork experience with the organization (and, in fact, in Kibera) and the rest were new hires. The canvassers were paid by the day, and wages were not tied to specific indicators of performance. Prior to the beginning of the door-to-door exercise, the canvassers were carefully trained on canvassing goals and methods. Teams were also given specific guidelines on the details of each treatment, along with a cheat sheet of the list of issues to be covered in each treatment group and a list of questions that were expected to frequently come up in the field.

Three layers of supervision were established over the sixty canvassers. The first layer was a team of seven "group leaders". Group leaders were randomly assigned four to five teams to accompany in the field. If teams fell behind schedule or if a team member was absent on a given day, the group leader was to step in and conduct canvassing visits instead. In practice, due to occasional absences and staff replacements, group leaders ended up being more involved in implementation than actual supervision or management. The second layer was a team of seven managers, all of whom were previous employees of the organization with more advanced fieldwork experience. The sole responsibility of the managers was to spend each day in the field, monitoring the teams under their supervision. They actively watched canvassers in the field and independently checked with a subset of households whether the canvassing visit had taken place as planned. Finally, the entire field exercise was supervised by a director of operations (a Kenyan national), a research assistant (foreign), and the principal investigators (i.e. the authors of this paper). Team managers reported back to the director of operations and the research assistant on a daily basis.

To maximize contact with households, each EA was covered twice during implementation. Given our previous work in Kibera, we expected that any member of a given household would be found at home a little less than half the time. Many household members spend the day outside their dwelling, either working or looking for employment. Activities were therefore planned so that each household would be visited twice over the two-week period, once each

week, and by a different team each week.

2.4 Diversity Experiment

Over the canvassing experiment, we collaborated with the organization to overlay an experiment on team diversity, with three layers of randomization. The canvassing exercise was conducted by teams of two. The canvassers were a mix of more experienced staff who had worked in the Kibera slum in 2012, and new staff hired specifically for the canvassing exercise. We randomly paired canvassers with each other (ensuring that no two experienced staff members were placed in the same team), and in doing so, introduced random variation in the degree of horizontal diversity within each team. Note that throughout our analysis, we show intent-to-treat estimates based on the initial random composition of teams, thus ignoring occasional staff replacements.

We then randomly allocated four to five teams to a manager, so as to induce random variation in vertical diversity. In the analysis below, "vertically diverse" teams are defined as teams where the ethnicity of the manager differs from that of both team members. Finally, we randomly allocated EAs (and hence households) to each team to create variation in external diversity, i.e. in how different teams were from the households on their canvassing list. Given the salience of ethnicity in Kenya, especially around elections, we were intesterested in testing whether canvassing outcomes differed in cases where the team members were of the same ethnicity as the household canvassed.

Each team was allocated to one of the two treatment groups, and assigned a random set of twenty EAs to cover (ten per week of canvassing). On average, any given team had to complete 1,045 visits over the two week period. The set of EAs assigned in the first week was different from the set of EAs assigned in the second week, so that a household would not be visited by the same canvassing team twice. The order in which canvassing teams visited the EAs allocated to them was also randomized. The rationale for this step of the randomization was to make sure that the experiment was uniformly implemented across the entire slum.

Canvassing teams used identifiable information (GPS coordinates and names of household heads) provided by the management to locate treatment households within the slum. Once they located a target household, canvassers were instructed to cover the relevant script for each treatment group, and to collect basic information about visits (this data is described in detail in Section 3). During training, canvassers were also encouraged to disseminate information to as many household/family members and/or neighbors as possible.[10]

[10]Since the treatment was allocated at the level of the EA, the staff was instructed to canvass everyone possible in the EA. In the tracking data, additional people are reported as being present in 1% of completed visits.

3 Data

3.1 Baseline and Randomization Data

The baseline data for this experiment came from two sources. First, we conducted a census of Kibera as part of prior work in early 2012 (Marx et al. (2014))). The census covered more than 30,000 households over two rounds of visits in the slum and was conducted using a very short survey module.[11] Second, also for the purpose of this earlier work, we were granted access to (de-identified) micro data for the entire slum from the 2009 national census, and to EA maps of the area. This allowed us to aggregate the 2009 census data to the EA level and to locate households visited in our 2012 census in specific EAs. Since the EAs were randomly allocated to the canvassing teams, we can test randomization balance by looking at correlations between staff characteristics and EA characteristics. We rely on our 2012 census, the 2009 national census, and related geospatial data for these balance checks (presented in section 5).

3.2 Performance Data

Each canvassing team was provided with tracking sheets that were pre-filled with information on households from the original 2012 census data. These tracking sheets included the names of target household heads, along with their GPS coordinates and village names. The canvassers were required to fill out information about each visit, in particular information on whether the household was found or not, the time when each visit began and ended, the number of household members present, and other visit details.[12]

In what follows, we use data from the tracking sheets to create two main measures of performance. First, we use a dummy variable for whether the canvassing visit was completed. This means the household was found and agreed to the canvassing visit (only 0.01% of the households successfully located by the staff refused to be canvassed). The second outcome is the duration of the visit. Based on the training given to canvassers, we expected a typical successful visit to last between five and ten minutes. In the data collected from the tracking sheets, successful visits lasted for 4.8 minutes on average. In our empirical analysis, we use an unconditional measure of time spent with each household, ranging from zero minutes (for unsuccessful visits) to an hour.[13] In addition to the untrimmed version of this variable, we compute a trimmed version where we drop values above the 99th percentile (10 minutes). Finally, we calculate the

[11] The questionnaire collected the name and ethnicity of the household head, household size, rental fees paid, phone numbers, and GPS coordinates.

[12] Canvassers were also required to fill out the name of the person in the household spoken to and their relationship to the household head; whether that person was already registered and whether they had a national ID card (needed to register); whether anyone else was present during the canvassing visit and if so, who they were.

[13] We treat as enumeration errors visits recorded to have lasted more than 60 minutes (these represent approximately 0.05% of the raw tracking data).

number of visits successfully conducted by each team on each day of fieldwork, the time spent in the field and the time spent canvassing, and we use these variables to corroborate our main results on effort and performance.

Although measures of performance are computed from the tracking sheets, we are not concerned about strategic misreporting by the canvassers driven by the demographic composition of teams. First, the organization's incentives structure was not tied to the number of successful canvassing visits or the duration of visits. Staff members were paid fixed daily wages regardless of the number of households successfully canvassed, and the staff was not given specific targets in terms of visit completion rates. Second, considerable effort and resources were spent towards monitoring and watching the teams in the field. For a total staff size of sixty canvassers, we had seven managers monitoring teams in the field on a daily basis, in addition to the director of operations and the research assistant. Third, gaming the measure of duration we use would have been difficult. The senior staff running the project could not credibly give daily feedback to canvassers based on the canvassing sheets, given the tracking sheets covered over 3,000 households each day – entering and analyzing the tracking sheets in real time was logistically infeasible.

Based on the tracking sheets, we also create a variable indicating whether teams conducted several visits simultaneously – effectively indicating that team members split to canvass households individually. Two visits are coded as simultaneous if they were conducted by the same team, on the same day, and completed within one minute of each other. We aggregate this measure by team and by day to compute the fraction of visits conducted individually in each team. This gives us a proxy for how different canvassing pairs organized their work on a daily basis.[14]

Finally, we use the data from the tracking sheets to create measures of effort. We compute daily measures of how much time the team spent in the field, how much time was spent canvassing, and how much time was spent simply walking between EAs and dwellings to find households (the difference between the first two measures). The time spent in the field was calculated as the difference between the start time of the first canvassing visit and the end time of last canvassing visit for each team and each day.[15] The time spent canvassing is the sum of the durations of all canvassing visits completed on a given day by each team.

Table 1 shows summary statistics from the performance data and the staff composition. Overall, 41% of all scheduled visits were successfully completed, in line with comparable experiments in the literature (see, for example, Pons and Liegey (2014)).[16] The unconditional visit duration averages 1.82 minutes. 23% of the canvassing pairs (7 teams out of 30) were ethnically

[14]Note this variable could also be capturing whether teams gathered households to conduct one visit with multiple households at the same time. This is consistent with our interpretation that the variable captures a more efficient organization of work within the team.

[15]We do not have a measure of the time spent searching for the first household every day.

[16]The mean success rate was 43.6% for horizontally ethnically homogeneous teams and 40.8% for horizontally ethnically diverse teams.

homogeneous, and 53% were either male-only or female-only teams. The ethnicity of manager matches that of one of the team members in 23% of cases. There is no instance where the ethnicity of the manager matches that of the two team members. Each team was allocated 1,045 visits on average, and managed to complete 426. On a daily basis, teams conducted 33 visits on average and spent approximately two and a half hours canvassing households – the rest of the time was spent walking between dwellings and searching for target households. Finally, approximately 14% of visits were conducted by one canvasser alone (as opposed to two).

3.3 Registration Data

Once the canvassing experiment was completed, we conducted a short phone call experiment in the final two days before the registration deadline. 8,100 households were randomly selected from an eligible pool to receive a phone call reminder about registration.[17] Households reached as part of this exercise were asked whether they had already registered for the upcoming election. 80% of respondents reported they had already registered at the time of the call. 58% reported they registered within Kibera, while 5% said they had no plans to register. We use these self-reported variables to measure final registration outcomes.

For the original canvassing experiment, we were also able to access the official voter register for the two constituencies (Kibra and Langata) that cover the Kibera area. Unfortunately, since we were not granted access to individual phone numbers or ID numbers, we are not able to confidently match participants in our experiment to the database of registered voters based on name only. To measure treatment effects on voter registration, we therefore rely on self-reported data collected from the phone calls described above.

3.4 Staff Survey Data

We conducted a survey of the canvassing staff 18 months after the completion of the experiment. We use data from this survey to support our interpretation of the main experimental results. The survey questionnaire was conducted individually and collected data on social interactions between teammates (e.g., did teammates interact socially after work or after the experiment), working methods (e.g. what language was primarily used within the team), as well as self-reported measures of working hours, and a self-assessment of their performance relative to that of their teammate. One of the canvassers could not be located at the time of this survey, bringing the number of respondents to fifty-nine.

[17] The 8,100 households selected represent (a random) half of households with a valid phone number in our 2012 census of Kibera. In the census, about 51% of the household heads reported a phone number. Unlike the canvassing experiment, this randomization was conducted at the household level. The phone calls were made by the same team of canvassers.

4 Empirical Framework

We now describe the empirical specifications we use to study the impact of team composition on two measures of team performance: a dummy for visits successfully completed, and the duration of each visit in minutes. For team composition, we focus primarily on the impact of ethnic homogeneity (horizontal and vertical), though we also discuss results for the impact of gender and age homogeneity.

The baseline specification is a regression of the following form:

$$y_{ijt} = \alpha + \beta EM_{jt} + \gamma^k_{jt} + \gamma^l_{jt} + \Omega X_{ij} + \epsilon_{ijt} \tag{1}$$

where y_{ijt} is a measure of the outcome of a canvassing visit for household i visited by team j in week t; EM_{jt} is a measure of horizontal ethnic homogeneity, i.e. a dummy equal to one if the two members of team j are of the same ethnic group; γ^k_{jt} and γ^l_{jt} are ethnicity dummies for each team member; and X_{ij} are a set of controls. The outcomes are indexed by time since each household was to be visited twice, so the data is at the household-week level. Throughout, we present intent-to-treat estimates, i.e. we use the value of the ethnic match variable EM_{jt} from the initial random allocation of teams.[18] In similar specifications, we also look at the effect of horizontal gender homogeneity and horizontal age homogeneity, where the latter is defined as the two junior teammates being within two years of age of each other.

In Table 3a, we first show specifications without controls, and then specifications where we control for the treatment group to which household i was allocated (this is just a dummy variable for the IEBC treatment), the (random) order in which household i was visited by team j in week t, and a dummy for the EA having more than median population (which was one of the variables we stratified the original treatment on). In our preferred specification, we cluster the standard errors by team-week. We show that our results are robust to alternative types of clustering, including a wild bootstrap clustering at the team level.

In addition to the tests of horizontal homogeneity, we ask whether the "vertical" composition of teams affect measured performance. There are no cases in the experiment where the ethnicity of the two team members and that of the manager all match. However, in 23% of cases the manager belongs to the same ethnic group as one of the team members. We examine whether this affects team performance, using the following specification:

$$y_{ijt} = \alpha + \beta_1 EM_{jt} + \beta_2 MEM_{jt} + \gamma^k_{jt} + \gamma^l_{jt} + \gamma^m_{jt} + \Omega X_{ij} + \epsilon_{ijt} \tag{2}$$

where MEM_{jt} is a dummy variable equal to one if the ethnicity of the manager matches the

[18] Note that we only use households allocated to either one of the two canvassing treatment groups to analyze the performance impacts of ethnic homogeneity. Since canvassing was only conducted in treatment EAs (and not in control EAs), outcome data is only available for households belonging to treatment EAs.

ethnicity of either team member; γ_{jt}^k, γ_{jt}^l and γ_{jt}^m are three sets of ethnicity dummies, one for each of the two junior team members and one for the manager. Note the correlation between the EM_{jt} and the MEM_{jt} dummies is -0.32 in our data. For comparison, we run a similar specification to test the effect of vertical diversity by gender, where vertical diversity means the gender of the manager differs from the gender of both junior canvassers.[19]

Finally, we study the effect of external homogeneity on performance, i.e. we ask whether the ethnicity of households randomly assigned to each team affects the performance of canvassers. For this, we use the following specification:

$$y_{ijt} = \alpha + \beta_1 EM_{jt} + \beta_2 HM_{ijt} + \beta_3 HMM_{ijt} + \gamma_{jt}^k + \gamma_{jt}^l + \gamma_{jt}^m + \gamma_i^n + \Omega X_{ij} + \epsilon_{ijt} \quad (3)$$

where HM_{ijt} is a dummy equal to 1 if the household has the same ethnicity as either team member, HMM_{ijt} is a dummy equal to 1 if the household has the same ethnicity as the manager, and γ_i^n are ethnicity dummies for household i.[20]

There are at least two distinct reasons why external homogeneity could affect our measures of performance. Canvassers could, in theory, target their efforts towards finding households of the same ethnicity if they feel a responsibility to mobilize co-ethnic voters in the upcoming election. Second, conditional on finding households, external homogeneity may also affect the time canvassers spend with respondents.

5 Results

5.1 Randomization Balance Checks

We first present a set of balance checks to verify the randomization. In particular, we check whether the measures of horizontal and vertical homogeneity correlate with any observable characteristics of the target areas and target households. Table 2 shows these results. Column (1) in this table reports the sample mean of the relevant dependent variable. In column (2), we show the coefficient obtained from regressing each relevant household and EA level characteristic on the horizontal homogeneity team dummy (equal to one for teams with co-ethnic canvassers) and ethnicity dummies for each team member. For EA-level variables, we run regressions at the EA-week level. In column (3), we report the coefficient from a similar specification using the measure of vertical homogeneity (coded as one if the manager is of the same ethnicity as any

[19] We do not look at a vertical age match as that is very unlikely to happen given tenure and experience patterns and given the experience needed to become a manager.

[20] We only consider external matches for households belonging to one of the seven following ethnic groups: Kalenjins, Kambas, Kikuyus, Kisiis, Luhyas, Luos, and Nubis (none of the staff members belong to the Nubi tribe, hence there are no external matches for Nubi households). Only 2% of households in the 2012 census data do not belong to any of these tribes. For these households, the external match dummy is unobserved (coded as missing).

one of the junior team members) as the main regressor. In Column (4), we report similar results for horizontal homogeneity by gender, and in column (5) for horizontal homogeneity by age. In column (6), we show results for the external match with the canvasser[21] All regressions include the relevant main effects.

The outcomes of these regressions are variables from the 2012 census as well as EA level aggregates of the 2009 census micro data. The former include household size, the number of children in the household, whether the household pays rent for their housing, the (log) amount paid in rent, whether the household was ever evicted from their dwelling in the past, the number of years spent in the same dwelling, the number of years spent in Kibera, and a dummy for households with a phone number in our data. The latter include log consumption per capita, the EA-level poverty rate,[22] the number of households in the EA (from the 2012 census), average age of the household head, years of education of the household head, whether the household head owns a business, a radio, and TV, whether the household head is employed, the fraction of youth unemployed in the EA, the average fraction of household dwellings with a cement floor, a mixed mud-cement wall, piped water, an uncovered pit latrine for sanitation, and electricity, and whether the household cook with paraffin.[23] The regressions using variables collected in Marx et al. (2014) are household-level regressions clustered by team-week, while the regressions using the 2009 census variables are EA-level regressions, clustered by team-week.

Table 2 illustrates that the randomization of the composition of canvassing teams produced a balanced experimental sample. Eight coefficients (out of 108) are significant at conventional levels (two at 10% and the remaining six at 5%).

5.2 Horizontal Ethnic Composition

The estimates from equation (1) are presented in Tables 3a and 3b for two outcomes: a dummy for successful visits, and the duration of canvassing visits. The main coefficient of interest is the coefficient on the horizontal ethnic match dummy. As a consequence of the experimental design, this coefficient captures the causal impact of a team's horizontal ethnic homogeneity on performance. We show estimates obtained from different specifications in Table 3a, and robustness to alternative clustering strategies in Table 3b.

The estimates shown in columns (1) and (2) of Table 3a indicate that ethnically homogeneous teams are 7 to 9 percentage points more likely to complete a canvassing visit. Such teams also conduct visits that are about 47% longer on average (columns (3)-(5)). The estimates in columns (1) and (3) are obtained from a version of specification (1) that includes no controls. In columns

[21] Here note that since the external match is defined at the household level, the only checks we can conduct are those with household characteristics from the 2012 census.

[22] The poverty indicator was computed as part of a poverty mapping exercise. See Marx et al. (2014) for details.

[23] These variables are standard measures of wealth in Sub-Saharan Africa. Some of them are used in the computation of the poverty indicator.

(2) and (4), we control for the IEBC Treatment, the order in which the canvassing team was instructed to visit that household, and a dummy for the EA having more than median population. In column (5), we show results for duration where we trim the top percentile of duration. The results are similar across these five columns. Note also that a position at the bottom of the canvassing order reduces the chance of a successful visit, and reduces visit duration.

In Table 3b, we show robustness to alternative ways of clustering our standard errors since there are different levels of experimental variation in our setup. Variation in the horizontal ethnic composition of teams is measured at the team level, and variation in the nature of visits (first vs. second visit) is at the week level. Our preferred clustering is therefore clustering by team-week, for which we have 60 clusters (30 teams times 2 weeks). Alternatively, we consider clustering by team only, in which case we also report the wild bootstrap p-value as there are only 30 clusters when we opt to cluster by team. Finally, we also show results from clustering by EA, since the main randomization (allocation of households to teams and allocation of households to an information treatment) was conducted at the EA level. As Table 3b shows, our main results are robust to these different ways of clustering our standard errors.

Columns (1) and (6) of Table 3b reproduce estimates from columns (2) and (4) in Table 3a, where standard errors are clustered by team-week. Standard errors are clustered by team in columns (2) and (7) and for these specifications, we also report the wild bootstrap p-value at the bottom of the table. In columns (3) and (8) we show results from clustering by EA. We use two-way clustering by team-week and by EA in columns (4) and (9), and two-way clustering by team and by EA in columns (5) and (10). Our results are robust across all these forms of clustering, including the wild bootstrap at the team level. Throughout, horizontal ethnic homogeneity has a strong positive effect on performance. We further show in Figure 2 that the fraction of co-ethnic pairs is a monotically increasing function of visit duration in the region containing most of the duration data (recall that 10 minutes is the 99th percentile of visit duration).

5.3 Vertical Ethnic Composition

Given that managers were randomly assigned to teams, we also exploit exogenous variation in a measure of vertical diversity, i.e. in whether the manager and one junior team member belong to the same ethnicity. Here we have no instances of a "triple match", i.e. cases where the manager and both members of the team belong to the same ethnicity. Thus "vertically homogeneous" teams in our context are teams where the manager belongs to the same ethnic group as one canvasser, and "vertically diverse teams" are teams where the manager's ethnicity differs from that of *both* canvassers.

Looking at both measures of team performance, we find evidence of effects that are opposite in sign to the effect of horizontal ethnic homogeneity (Table 4). Remembering that this effect is identified from canvassing pairs that are not ethnically homogeneous themselves, the mag-

nitudes are sizeable. In particular, if a manager and any one team member belong to the same ethnicity, the probability that the canvassing visit is completed decreases by about 6 percentage points (though this is not statistically significant) and the duration of visits decreases considerably. The effects on duration are sizeable and statistically significant, even when we cluster by team and rely on wild bootstrap p-values for inference. In addition, the treatment effects are statistically identical when we disaggregate vertical matches into two cases: the ethnic match is between the manager and the rookie canvasser, or between the manager and the experienced canvasser (columns (2) and (6)). In Figure 3, we also show that vertical homogeneity is a monotically decreasing function of visit duration across all visits lasting less than 10 minutes (the 99th percentile).

5.4 Direct Effect of Ethnic Composition on Visit Duration

Throughout our analysis, we focus on an unconditional measure of visit duration, coding the duration of unsuccessful visits as zero, since the measured duration of successful visits is only available for a selected subset of households (and visit success is itself an outcome of ethnic composition). To address the endogeneity of visit completion and disentangle the effect on completion versus that on duration, in Table 5 we estimate a Heckman selection model for conditional visit duration, i.e. where visit duration is the outcome but is only observed for the sample of households where a visit was completed.

The variable included in the first stage probit, but excluded from the outcome equation, is a dummy variable indicating whether a household was successfully located twice (i.e. in both phases) in the initial census of the slum conducted in 2012 (see Section 3.1). This was the case for 31% of households in the census data. We expect households found twice in the census to be more likely to be successfully canvassed in the canvassing experiment, since having been found twice suggests a household member was present at all times in the dwelling or the household was well-known to the neighborhood. The bottom panel of Table 5 suggests this variable is indeed a strong positive predictor of visit completion. In addition, we think that this variable, whether a household was found twice during the census in 2012, would not affect the duration of the canvassing visits directly, just whether a visit was completed. The top panel of Table 5 shows that after addressing the endogeneity of visit completion, horizontal ethnic homogeneity still positively affects visit duration, while vertical homogeneity has the opposite effect. These findings are consistent with the estimates presented in Tables 3a, 3b and 4.

5.5 Homogeneity by Gender and Age

In Table 6, we look at the performance effects of team homogeneity by gender and age. Standard errors are reported using the same three clustering strategies: by team-week, by team and by EA.

We do find that gender-homogeneous teams perform better (columns (1)-(6)), but largely for the completion of canvassing visits (although the magnitudes are smaller) and not for the duration of the visits. On completion, we find that teams homogeneous along gender lines are about 4 percentage points more likely to complete a canvassing visit. Visits are also 6% longer with a gender match, though this effect is not statistically different from zero. Although these effects are economically sizeable, it is worth noting that they are much smaller than the effects of ethnic homogeneity reported in Table 3a – about half the size for completion, and one eighth the size for duration. In columns (3) and (6), we look at vertical diversity along the gender dimension. Vertical gender diversity decreases the rate of visit completion (only marginally significant) but it has no effect on visit duration.

Columns (7)-(10) of Table 6 report effects of team homogeneity by age. We find that homogeneity by age significantly increases the chance of a successful canvassing visit, but has no effect on visit duration. As with homogeneity by gender, the effect of age homogeneity on the success of visits is much smaller in magnitude than the effect of ethnic homogeneity.

5.6 Ethnic matches with households

In Table 7, we look at the impact of external diversity, i.e. whether the households' ethnicity affects the success and duration of canvassing visits (equation (3)). The effect of an external ethnic match is a precisely estimated zero – we are able to rule out very small effects. There is also no effect of an external ethnic match with the manager. We do not report effects of a "triple match" between the household's tribe and both canvassers on the team, since this configuration occurs in only 3% of all cases. This effect is negative, small in magnitude, and not significantly different from zero (results available upon request).

5.7 Effects on Registration

In Table 8, we look at the impacts of the horizontal ethnic homogeneity of canvassing teams on self-reported measures of household registration. Since this measure is only available for households we could reach by phone in the last days before the registration deadline, the sample size drops to less than 7,000 households (the response rate in our phone survey was 84%).

For comparison, in columns (1) and (2) we reproduce our estimates of the effect of horizontal ethnic homogeneity in this limited sample. In this limited sample, ethnic homogeneity within the team of canvassers still positively affects the completion and duration of visits, and the effects are of the same magnitude as in Table 3a. There is no such positive effect when measures of registration are used as dependent variables. The four outcomes we look at are whether the head of household reported being registered (column (3)), whether the entire household registered (column (4)), whether the head of household has made no plans to register (column (5)),

and whether the head of household registered in a Kibera polling station (column (6)). The horizontal ethnic match dummy has no effect on any of these variables. Preliminary analysis of the overall registration experiment exploiting both the self-reported data as well as the administrative registration data suggest that the canvassing intervention simply did not significantly increase voter registration (results not reported).

6 Mechanisms and Discussion

To better understand mechanisms, we present two sets of further results: the first set focuses on measures of team organization, effort and productivity, and the second on outcomes collected from our survey of the canvassers.

We show evidence on the organization of work within teams in Table 9. Our key outcome of interest in this table is whether teams split to conduct canvassing visits individually. In columns (1)-(4), we regress a dummy variable for visits conducted simultaneously (defined as in Section 3.2) on the usual set of covariates, including dummies for horizontally homogeneous and vertically homogeneous teams. We find that members of a horizontally homogeneous team are significantly more likely to complete any visit individually, while vertically homogeneous teams are no more likely to do so. The model in columns (1)-(2) is estimated via OLS (as in Tables 3a, 3b, and 4), while in columns (3)-(4) we present estimates from a Heckman selection model analogous to the model presented in Table 5 (where visit completion is the dependent variable in the selection equation).

In columns (5)-(8) we present findings on the number of successful visits conducted each day, and the fraction of visits conducted individually each day, based on a dataset organized at the team-day level. We find that horizontally homogeneous teams complete a larger number of visits per day (columns (5)-(6)) and a larger fraction of visits individually (columns (7)-(8)). Vertical homogeneity has no statistically significant effect on either outcome. While we did not give instructions on splitting teams during training, this seems to have been a productive decision as it allowed horizontally homogeneous teams to complete more visits overall.[24]

In Table 10 and Figure 4, we present further results on effort and productivity. We focus on three outcomes: how much time each team spent on a given day in the field; how much time they spent actually canvassing; and how much time they spent walking between dwellings or searching for households, which is simply the difference between the first two measures. We interpret time spent in the field as a measure of effort and the time split between canvassing and non-canvassing as a measure of team productivity or efficiency. Figure 4 shows the dis-

[24]The effect of splitting the team on performance is ex ante ambiguous. On the one hand, keeping the team together at all times might have promoted synergies and learning. On the other hand, splitting visits obviously allowed the team to conduct a larger number of visits.

tributions of these measures, separately for teams that are ethnically homogeneous (along the horizontal dimension) and those that are not. Ethnically homogeneous teams spend about the same amount of time in the field, but a much larger amount of time canvassing than diverse teams. Table 10 supports these results. The teams that match horizontally do not exert more effort (time spent in the field does not differ between homogeneous and diverse teams), but they spend more time canvassing and significantly less time walking or searching. Teams that have a vertical ethnic match with the manager spend both less time in the field and less time canvassing. This is evidence that effort levels are lower within those teams.

In Table 11, we present results from the survey conducted with the canvassing staff. This survey covered a number of aspects of the team exercise, including team communication, whether canvassers socialized with their teammates after the experiment, measures of how they worked together as a team, and self-reported measures of effort and performance. We regress these outcomes on the horizontal and the vertical ethnic match dummies, as well as tribe main effects.

In columns (1) and (2), we look at two self-reported measures of performance. The first measure asked the canvasser to rank his own performance on a scale of 1 to 5 (from very bad to very good). In column (1), we report results for a dummy variable capturing whether the canvasser ranked his performance as very good. We find that teams with a vertical ethnicity match report lower own performance. The second measure of performance was relative: we asked canvassers to report whether they performed better than their teammate or vice versa. In column (2), we report results for this variable. Canvassers belonging to an ethnically homogeneous team along the horizontal dimension were significantly less likely to say they performed better than their teammate (as were teams that had a vertical ethnic match, though not significantly so).

Similarly, in column (3), we show that horizontally homogeneous teams were significantly less likely to complain about their teammate, as were vertically homogeneous teams. In column (4) we look at a self-reported measure of effort. Hours worked are no different in horizontally homogeneous teams, but are significantly lower for vertically matched teams (both results are consistent with those in Table 10). In column (6), we look at the time the team spent brainstorming. We find no differences across homogeneous and diverse teams along this metric.

Based on our main results and on the survey data, preferences towards working with similar workers are unlikely to explain our results for two reasons. First, canvassers are no more likely to spend time with households of the same ethnicity, as shown in Table 7. Second, we find that teammates of the same ethnicity were no more likely to socialize during or after the exercise, or to still be in contact a year later (results not reported). We also do not find that teams of the same ethnicity used their tribal language as their primary working language – only one team out of thirty did so, while all others communicated in either English or Swahili, the main language spoken throughout Kenya.

Overall, we interpret these results as evidence that the horizontal homogeneity within teams

acted as a monitoring and disciplining device and improved work organization within the team. Meanwhile, vertical homogeneity led to less stringent norms and discipline and hence lower effort and performance. This finding is consistent with Bandiera et al. (2009), who show that an increase in the average level of social connections between managers and workers in a fruit-picking company can reduce average firm productivity by inducing distortions in managerial effort (under fixed wages).[25] These results highlight the importance of looking at different dimensions of diversity when analyzing the effect of diversity on performance. There may be costs to diversity in terms of the horizontal structure of the firm – but these costs must be balanced against the large gains of diversity we find along the vertical dimension.

7 Conclusion

In this paper, we study the effects of team diversity on performance. We use a field experiment implemented within a non-profit research organization based in Kenya, as part of a door-to-door canvassing exercise where pairs of canvassers were assigned a list of households to visit. We use the data from this canvassing exercise to compute measures of effort and performance of the staff. Our most innovative contribution is that we are able to study the causal effects of diversity along various dimensions (within teams, along the hierarchy of the organization, and between workers and clients) within a single organization. We also conducted a survey of the staff that sheds light of the mechanisms underlying the effects of team diversity on effort and performance.

We find that ethnic homogeneity between teammates (the horizontal dimension) improves team performance, measured in terms of the completion and duration of each canvassing visit assigned to the team. The magnitude of the effect of ethnic homogeneity is economically sizeable – for example, ethnically homogeneous teams are about 9% more likely to successfully complete any canvassing visit. This effect is more than twice the size of the effect of homogeneity by gender or age group. Ethnic homogeneity between one junior team member and the manager (the vertical dimension) has the opposite effect: teams with this configuration report poorer performance on average. Finally, we find no effect of an ethnic match between households and canvassing teams or managers. This null result suggests that teams did not strategically direct their efforts towards co-ethnic households.

Our analysis of time use suggests that while horizontally homogeneous teams do not spend more time in the field, they organize their time more efficiently and are able to spend more time actually canvassing households. Conversely, vertical ethnic homogeneity reduces both the time spent in the field and the time spent canvassing, suggesting a lower level of effort (and performance) in these teams. We interpret these results in light of evidence from a survey of the

[25] The introduction of managerial incentives can undo this effect, but such incentives were absent in our context.

canvassing staff. First, the survey helps us rule out simple preference-based or language explanations for the different performances of teams. Ethnically homogeneous teams do not report using their tribal language as their primary working language, and they do not report more social interactions with teammates of the same ethnicity during or after the exercise. However, there is evidence that horizontally homogeneous teams are more cohesive and more likely to divide up tasks. Canvassers in vertically homogeneous teams report working (significantly) less hours and provide a more negative assessment of their own performance.

Our findings suggest much of the trade-off between diversity and homogeneity in organizations may come from the different effects diversity has along different dimensions of organizational structure. On the one hand, diversity may reduce efficiency within teams of workers, by creating communication costs or other frictions, leading to a worse division of tasks and lower performance. On the other hand, diversity along the organization's hierarchy has the opposite effect in our context, since it improves both effort and performance. A relevant direction for future research could be to look at the reasons why diversity along the hierarchy induces more effort and/or different norms within teams.

References

Aker, Jenny C., Paul Collier, and Pedro C. Vicente, "Is Information Power? Using Cell Phones During an Election in Mozambique," *Working Paper*, 2011.

Alesina, Alberto, Reza Baqir, and William Easterly, "Public Goods and Ethnic Divisions," *Quarterly Journal of Economics*, 1999, *114*, 1243–1284.

Alesina, Roberto and Eliana La Ferrara, "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, 2005, *43* (3), 762–800.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson, "The Impact of Jury Race in Criminal Trials," *Quarterly Journal of Economics*, 2012, *127* (2), 1017–1055.

Apfelbaum, Evan P., Katherine W. Phillips, and Jennifer A. Richeson, "Rethinking the Baseline in Diversity Research: Should We Be Explaining the Effects of Homogeneity," *Perspectives on Psychological Science*, 2014, *9* (3), 235–244.

Arrow, Kenneth, "The Theory of Discrimination," in O. Ashenfelter and A. Rees, eds., *Discrimination in Labor Markets*, Princeton, NJ: Princeton University Press, 1973.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul, "Social Connections and Incentives in the Workplace: Evidence From Personnel Data," *Econometrica*, 2009, *77* (4), 1047–1094.

Banerjee, Abhjit V., Selvan Kumar, Rohini Pande, and Felix Su, "Do Informed Voters Make Better Choices? Experimental Evidence from Urban India," *Working Paper*, 2010.

Becker, Gary, *The Economics of Discrimination*, University of Chicago Press, 1957.

Braconnier, Céline, Jean-Yves Dormagen, and Vincent Pons, "Voter Registration Costs and Disenfranchisement: Experimental Evidence from France," *Working Paper*, 2014.

Easterly, William and Ross Levine, "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics*, 1997, *112* (4), 1203–1250.

Fafchamps, Marcel, "Ethnicity and Credit in African Manufacturing," *Journal of Development Economics*, 2000, *61*, 205–235.

Fairlie, Robert W., Florian Hoffmann, and Philip Oreopoulos, "A Community College Instructor Like Me: Race and Ethnicity Interactions in the Classroom," *American Economic Review*, 2014, *104* (8), 2567–2591.

Fisman, Raymond, Daniel Paravisini, and Vikrant Vig, "Cultural Proximity and Loan Outcomes," *NBER Working Paper No. 18096*, 2012.

Hamilton, Barton H., Jack A. Nickerson, and Hideo Owan, "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation," *Journal of Political Economy*, 2003, *111* (3), 465–497.

Hamilton, Barton H, Jack A Nickerson, and Hideo Owan, "Diversity and Productivity in Production Teams," *Advances in the Economic Analysis of Participatory and Labor-Managed Firms*, 2012, *13*, 99–138.

Hjort, Jonas, "Ethnic Divisions and Production in Firms," *Quarterly Journal of Economics*, 2014, *129* (4).

Hoogendoorn, Sander and Mirjam van Praag, "Ethnic Diversity and Team Performance: A Field Experiment," *Tinbergen Institute Discussion Paper 2012-068/3*, 2014.

_ , Simon C. Parker, and Mirjam van Praag, "Ability Dispersion and Team Performance: A Field Experiment," *Tinbergen Institute Discussion Paper 2012-130/VII*, 2012.

Kandel, Eugene and Edward P. Lazear, "Peer Pressure and Partnerships," *Journal of Political Economy*, 1992, *100* (4), 801–817.

Katherine, S.Y. Kim-Jun Phillips and S. Shim, "The Value of Diversity in Prganizations: A Social Psychological Perspective," in R. van Dick David De Cremer and K. Murnighan, eds., *Social Psychology and Organizations*, New York, NY: Routledge, 2011.

Lang, Kevin, "A Language Theory of Discrimination," *The Quarterly Journal of Economics*, 1986, *101* (2), 363–382.

Lazear, Edward P., "Globalisation and the Market for Team-Mates," *The Economic Journal*, 1999, *109*, C15–C40.

_ , Kathryn L. Shaw, and Christopher T. Stanton, "The Value of Bosses," *forthcoming Journal of Labor Economics*, 2014.

Levine, Sheen S., Evan P. Apfelbaum, Mark Bernard, Valerie L. Bartelt, Edward J. Zapac, and David Stark, "Ethnic diversity deflates price bubbles," *PNAS*, 2014, *111* (52), 18524–18529.

Lyons, Elizabeth, "Team Production in International Labor Markets: Experimental Evidence from the Field," *Working Paper*, 2014.

Marx, Benjamin, Thomas Stoker, and Tavneet Suri, "There Is No Free House: Ethnic Patronage and Diversity in a Kenyan Slum," *Working Paper*, 2014.

Mas, Alexandre and Enrico Moretti, "Peers at Work," *American Economic Review*, 2009, *99* (1), 112–145.

Miguel, Edward and Mary Kay Gugerty, "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya," *Journal of Public Economics*, 2005, *89*, 2325–2368.

Papps, Kerry L., Alex Bryson, and Rafael Gomez, "Heterogeneous worker ability and team-based production: Evidence from major league baseball, 19202009," *Labour Economics*, 2011, *18* (3).

Pons, Vincent and Guillaume Liegey, "Increasing the Electoral Participation of Immigrants: Experimental Evidence from France," *Working Paper*, 2014.

Prat, Andrea, "Should a Team be Homogeneous?," *European Economic Review*, 2002, *46*, 1187–1207.

Price, Joseph and Justin Wolfers, "Racial Discrimination among NBA Referees," *The Quarterly Journal of Economics*, 2010, *125* (4), 1859–1887.

Reinhard, Selten and Massimon Warglien, "The Emergence of Simple Languages in an Experimental Coordinate Game," *PNAS*, 2007, *104* (18), 7361–7366.

Shayo, Moses and Asaf Zussman, "Judicial Ingroup Bias in the Shadow of Terrorism," *Quarterly Journal of Economics*, 2011, *126*, 1447–1484.

Zenger, Todd R. and Barbara S. Lawrence, "Organizational Demography: The Differential Effects of Age and Tenure Distributions on Technical Communication," *Academy of Management Journal*, 1989, *32* (2), 353–376.

Table 1: Summary Statistics

	Mean	SD	*N*
Household-Week Level Data			
Visit Completed	0.41	0.49	30947
Duration of Visit	1.82	3.02	29302
Duration of Visit, Trimmed	1.67	2.48	29017
Duration, Conditional on Finding Household	4.79	3.13	11135
External Ethnic Match: Canvasser	0.30	0.46	27543
External Ethnic Match: Manager	0.16	0.37	27543
Household Found Twice in Census (instrument)	0.31	0.46	30947
Visit was Conducted Alone (1 canvasser)	0.05	0.22	30946
Team-Level Data			
Total Visits Allocated to Team	1045.07	117.72	30
Total Visits Completed by Team	426.00	73.27	30
Horizontal Ethnic Match	0.23	0.43	30
Vertical Ethnic Match	0.23	0.43	30
Horizontal Gender Match	0.53	0.51	30
Vertical Gender Match	0.67	0.48	30
Horizontal Age Match	0.47	0.51	30
Total EAs Allocated to Team	20.00	0.37	30
Team-Day Level Data			
# of Visits Conducted	32.85	18.32	279
% of Alone Visits Conducted	0.14	0.09	279
Time Spent Canvassing	155.97	97.88	279
Time Spent in Field	319.41	132.78	274
Idle Minutes Spent in the Field	169.62	103.77	266

Note: Each of the 30 teams was composed of two staff members.
Each team covered 10 EAs each week for a total of two weeks.
Each household was visited twice, each time by a different team.
Duration data is missing for 1,645 completed visits.
The external match is missing when household ethnicity is unobserved.

Table 2: Randomization Checks

	Mean	Horizontal Ethnic Match	Vertical Ethnic Match	Horizontal Gender Match	Horizontal Age Match	External Match
Household Size	3.6692	.0029	.0542	-.0439	-.0638	-.0568
		[.0538]	[.0888]	[.0393]	[.0421]	[.0354]
# Children in HH	1.5524	.005	-.0119	-.0404	-.0454	-.0265
		[.045]	[.0835]	[.035]	[.0342]	[.0271]
Household pays rent	.9231	-.0093	-.0097	-.0042	-.0068	.0014
		[.0076]	[.0176]	[.0051]	[.0062]	[.0061]
Housing Rent (log)	7.2695	-.0498	.0699	-.0383	.0148	-.0158
		[.0553]	[.0877]	[.0375]	[.034]	[.0178]
Ever evicted	.1222	.0007	.0054	.0032	-.007	.0079
		[.005]	[.0127]	[.0063]	[.0066]	[.0067]
Years in Structure	8.2269	.196	.1559	.2528	-.0203	.0381
		[.2543]	[.8406]	[.1892]	[.2238]	[.1911]
Years in Kibera	15.5485	.1173	-.087	.2822	-.0487	-.0292
		[.3555]	[1.1647]	[.2389]	[.3125]	[.1994]
Household Has Phone	.5125	.0299	.0257	.0029	-.0344**	-.0127
		[.0195]	[.0375]	[.0126]	[.0148]	[.0085]
Log Consumption p.c.	10.4168	.0061	.0456	-.0062	.0029	
	.	[.0215]	[.0505]	[.0154]	[.0171]	
EA Poverty Rate	.1467	-.0053	-.0273	-.0044	-.0034	
		[.0096]	[.0248]	[.0062]	[.0073]	
# Households in EA	52.2533	.9767	1.6603	-4.857**	-3.6712	
		[2.7272]	[7.5348]	[1.8793]	[2.2617]	
Age of HH Head	35.4021	-.0496	-.0245	.0928	-.0462	
		[.2741]	[.7055]	[.1689]	[.2133]	
Years of Education	9.2715	-.0062	.4953**	-.1093	-.0428	
		[.1711]	[.2234]	[.1112]	[.1233]	
Owns a Business	.1496	-.016	-.0122	.0138	.0002	
		[.0134]	[.0338]	[.0094]	[.0108]	
Unemployed	.3157	-.0194	-.0103	.0207**	.003	
		[.0161]	[.0349]	[.01]	[.0102]	
% Youth Unemployed	.1682	.0146	.0143	.0059	.0059	
		[.013]	[.0314]	[.0114]	[.0115]	
Own Radio	.7542	.0096	.0346	.0066	.0066	
		[.0149]	[.0367]	[.0104]	[.0125]	
Own TV	.4143	.0305	.0149	-.0158	-.0221	
		[.023]	[.0734]	[.0165]	[.0226]	
Mud Cement Walls	.511	.0551	.0016	-.0018	-.0484*	
		[.0375]	[.0859]	[.0275]	[.028]	
Cement Floor	.6331	.0392	.0592	-.0025	-.032	
		[.0337]	[.0826]	[.0216]	[.0266]	
Piped Water	.6813	.0571	-.1657	-.053	-.0952**	
		[.0537]	[.1384]	[.0366]	[.0355]	
Covered Pit	.6582	.1076**	-.2077**	.0159	-.0428	
		[.0467]	[.0903]	[.0314]	[.0333]	
Cook with Paraffin	.5325	.0389	.0108	.0141	.0121	
		[.0237]	[.0351]	[.0142]	[.0152]	
Electricity	.5135	.0331	-.0086	-.0301	-.0475	
		[.0323]	[.0961]	[.0239]	[.0303]	

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team-week in brackets.
HH stands for Household. Column (1) shows the mean of each row variable.
Each cell shows estimates from a regression of the row variable on the column variable and relevant main effects.
In col. (6), the match is defined at the household level hence only results for household level variables are reported.

Table 3a: Horizontal Ethnic Composition and Team Performance

	(1) Visit Completed	(2) Visit Completed	(3) Visit Duration	(4) Visit Duration	(5) Visit Duration (trimmed)
Horizontal Ethnic Match	0.067***	0.092***	0.866***	0.855***	0.740***
	[0.023]	[0.027]	[0.209]	[0.220]	[0.209]
Random Visit Order		-0.004***		-0.032***	-0.030***
		[0.001]		[0.009]	[0.008]
IEBC Treatment		-0.048**		0.032	-0.056
		[0.020]		[0.158]	[0.152]
EA population		-0.008		-0.038	-0.059
		[0.009]		[0.053]	[0.047]
R^2	0.008	0.011	0.024	0.027	0.031
Clustering by Team-Week	X	X	X	X	X
Controls		X		X	X
Clusters	60	60	60	60	60
Dep Var Mean	0.413	0.413	1.821	1.821	1.000
Observations	30947	30947	29302	29302	29017

Note: * p<0.1, ** p<0.05, *** p<0.01.

All specifications include ethnicity dummies for each staff member.

In column (5), we drop the top percentile in visit duration (visits longer than 10 minutes).

Table 3b: Horizontal Ethnic Composition and Team Performance, Robustness to Clustering

	(1) Completed	(2) Completed	(3) Completed	(4) Completed	(5) Completed	(6) Duration	(7) Duration	(8) Duration	(9) Duration	(10) Duration
Horizontal Ethnic Match	0.092***	0.092***	0.092***	0.092***	0.092***	0.855***	0.855***	0.855***	0.855***	0.855***
	[0.027]	[0.030]	[0.016]	[0.026]	[0.028]	[0.220]	[0.276]	[0.110]	[0.213]	[0.267]
Random Visit Order	-0.004***	-0.004***	-0.004***	-0.004***	-0.004***	-0.032***	-0.032***	-0.032***	-0.032***	-0.032***
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.009]	[0.008]	[0.006]	[0.009]	[0.008]
IEBC Treatment	-0.048**	-0.048**	-0.048***	-0.048**	-0.048**	0.032	0.032	0.032	0.032	0.032
	[0.020]	[0.020]	[0.015]	[0.021]	[0.020]	[0.158]	[0.195]	[0.082]	[0.153]	[0.189]
EA population	-0.008	-0.008	-0.008	-0.008	-0.008	-0.038	-0.038	-0.038	-0.038	-0.038
	[0.009]	[0.009]	[0.012]	[0.010]	[0.010]	[0.053]	[0.057]	[0.065]	[0.052]	[0.056]
R^2	0.011	0.011	0.011	0.011	0.011	0.027	0.027	0.027	0.027	0.027
Clustering by Team-Week	X			X		X			X	
Clustering by Team		X			X		X			X
Clustering by EA			X	X	X			X	X	X
Clusters	60	30	300			60	30	300		
Wild Bootstrap P-value		0.006					0.012			
Dep Var Mean	0.413	0.413	0.413	0.413	0.413	1.821	1.821	1.821	1.821	1.821
Observations	30947	30947	30947	30947	30947	29302	29302	29302	29302	29302

Note: * p<0.1, ** p<0.05, *** p<0.01.

All specifications include ethnicity dummies for each staff member.

Results for duration are unchanged if we trim the top percentile of visit durations.

Table 4: Vertical Ethnic Composition and Team Performance

	(1) Completed	(2) Completed	(3) Completed	(4) Completed	(5) Duration	(6) Duration	(7) Duration	(8) Duration
Vertical Ethnic Match	-0.058		-0.058	-0.058	-1.251***		-1.251***	-1.251***
	[0.038]		[0.045]	[0.039]	[0.278]		[0.329]	[0.227]
Vertical Match: Experienced		-0.055				-1.092***		
		[0.039]				[0.298]		
Vertical Match: Rookie		-0.062				-1.558***		
		[0.041]				[0.327]		
Horizontal Ethnic Match	0.079***	0.079***	0.079***	0.079***	0.787***	0.754***	0.787***	0.787***
	[0.029]	[0.029]	[0.022]	[0.021]	[0.214]	[0.211]	[0.237]	[0.129]
R^2	0.013	0.013	0.013	0.013	0.030	0.031	0.030	0.030
Clustering by Team-Week	X	X			X	X		
Clustering by Team			X				X	
Clustering by EA				X				X
Clusters	60	60	30	300	60	60	30	300
Wild BS P-val: Vertical			0.390				0.002	
Wild BS P-val: Horizontal			0.060				0.042	
Dep Var Mean	0.413	0.413	0.413	0.413	1.821	1.821	1.821	1.821
Observations	30947	30947	30947	30947	29302	29302	29302	29302

Note: * p<0.1, ** p<0.05, *** p<0.01.

All specifications include ethnicity dummies for each staff member (including the team manager).

All specifications include controls for IEBC treatment, random order and EA population.

Table 5: Effect on Visit Duration, Heckman Correction

	(1) Duration	(2) Duration	(3) Duration	(4) Duration
Duration, Conditional on Finding Household				
Horizontal Ethnic Match	1.563***	1.541***	1.337***	1.376***
	[0.430]	[0.443]	[0.370]	[0.358]
Vertical Ethnic Match			-2.289***	-2.898***
			[0.625]	[0.549]
Random Visit Order		-0.067***		-0.067***
		[0.018]		[0.017]
IEBC Treatment		0.094		-0.662
		[0.355]		[0.516]
EA population		-0.062		-0.054
		[0.104]		[0.099]
Selection Equation (Household Canvassed)				
Household found twice in census (instrument)	0.103***	0.102***	0.102***	0.102***
	[0.012]	[0.012]	[0.012]	[0.013]
Horizontal Ethnic Match	0.281***	0.326***	0.220**	0.235**
	[0.083]	[0.093]	[0.096]	[0.096]
Vertical Ethnic Match			0.154	-0.163
			[0.121]	[0.112]
Random Visit Order		-0.011***		-0.011***
		[0.003]		[0.003]
IEBC Treatment		-0.085		-0.326***
		[0.068]		[0.078]
EA population		-0.015		-0.021
		[0.022]		[0.021]
Dep var mean	4.792	4.792	4.792	4.792
Controls		X		X
Inverse Mills ratio	4.5	4.5	4.4	4.4
S.E. Inv. Mills	[0.2]	[0.2]	[0.2]	[0.2]
Observations	29302	29302	29302	29302

Note: * p<0.1, ** p<0.05, *** p<0.01.

The dependent variable is conditional visit duration (average: 4.8 minutes).

The excluded variable used to predict visit completion is a dummy variable indicating whether the household was found twice during the 2012 census.

Table 6: Gender and Age Composition and Team Performance

	Completed			Duration			Completed		Duration	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Vertical Gender Match			0.042*			-0.042				
			[0.023]			[0.180]				
Horizontal Gender Match	0.037**	0.037*	0.062***	0.116	0.116	0.117				
	[0.017]	[0.021]	[0.018]	[0.139]	[0.184]	[0.140]				
Horizontal Age Match							0.035*	0.035	-0.044	-0.044
							[0.019]	[0.023]	[0.183]	[0.240]
R^2	0.005	0.005	0.007	0.009	0.009	0.011	0.008	0.008	0.011	0.011
Clustering by Team-Week	X		X	X		X	X		X	
Clustering by Team		X			X			X		X
Wild Bootstrap P-value		0.124			0.545			0.154		0.971
Clusters	60	30	60	60	30	60	58	29	58	29
Dep Var Mean	0.413	0.413	0.413	1.821	1.821	1.821	0.414	0.414	1.805	1.805
Observations	30947	30947	30947	29302	29302	29302	30156	30156	28544	28544

Note: * p<0.1, ** p<0.05, *** p<0.01.

All regressions include main effect dummies, i.e. canvasser-level controls for gender and age, as needed.

A dummy for the IEBC treatment, the random visit order and a dummy for large EA population are also included in all regressions.

The vertical gender match is a dummy for whether the manager has the same gender as either one of the two junior staff members.

The age match is a dummy for whether the two staff members are two or less years apart in age.

Age is missing for one team, hence the lower sample sizes (and one fewer clusters) for the age match regressions.

Table 7: Ethnicity of the Households and Team Performance

	(1) Completed	(2) Completed	(3) Completed	(4) Duration	(5) Duration	(6) Duration
External Ethnic Match: Canvasser	0.005	0.005	0.005	-0.008	-0.008	-0.008
	[0.008]	[0.008]	[0.009]	[0.055]	[0.054]	[0.057]
External Ethnic Match: Manager	0.004	0.004	0.004	-0.044	-0.044	-0.044
	[0.010]	[0.011]	[0.012]	[0.084]	[0.087]	[0.079]
Horizontal Ethnic Match	0.084***	0.084**	0.084***	0.810***	0.810***	0.810***
	[0.022]	[0.035]	[0.020]	[0.260]	[0.248]	[0.128]
R^2	0.022	0.022	0.022	0.032	0.032	0.032
Clustering by Team	X			X		
Clustering by Team-Week		X			X	
Clustering by EA			X			X
Clusters	30	60	299	30	60	299
Dep var mean	0.422	0.422	0.422	1.866	1.866	1.866
Observations	27543	27543	27543	26067	26067	26067

Note: * p<0.1, ** p<0.05, *** p<0.01.

All specifications include ethnicity dummies for each staff member and household tribe dummies.

All specifications include controls for IEBC treatment, random order and EA population.

Household ethnicity data is unavailable for 1,376 households in our sample.

Table 8: Self-Reported Registration Outcomes

	(1) Completed	(2) Duration	(3) Registered	(4) Family Registered	(5) No Registration Plans	(6) Registered in Kibera
Horizontal Ethnic Match	0.075**	0.892***	-0.016	-0.004	0.008	-0.005
	[0.033]	[0.253]	[0.012]	[0.017]	[0.010]	[0.018]
R^2	0.008	0.032	0.002	0.002	0.002	0.004
Clusters	60	60	60	60	60	60
Dep var mean	0.488	2.159	0.799	0.686	0.050	0.585
Observations	6721	6313	6782	6678	6774	6782

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team-week in brackets.
All specifications include ethnicity dummies for each staff member.
Registered (column 3) indicates whether the head of household registered.
Family Registered (column 4) indicates whether the entire household registered.
No Registration Plans (column 5) indicates the head of household had no plans to register.
Registered in Kibera (column 6) indicates the head of household registered to vote in Kibera.

Table 9: Visits Conducted Alone

	(1) Visits Alone	(2) Visits Alone	(3) Visits Alone	(4) Visits Alone	(5) # Visits	(6) # Visits	(7) % Visits Alone	(8) % Visits Alone
Horizontal Ethnic Match	0.040***	0.031***	0.065***	0.050***	7.432	10.545***	0.063**	0.048*
	[0.010]	[0.010]	[0.017]	[0.016]	[4.530]	[3.448]	[0.024]	[0.026]
Vertical Ethnic Match		-0.017		-0.011		4.806		0.064
		[0.017]		[0.033]		[5.880]		[0.039]
R^2	0.007	0.009			0.075	0.122	0.089	0.151
Model	OLS	OLS	Heckman	Heckman	OLS	OLS	OLS	OLS
Controls	X	X	X	X	X	X	X	X
Clusters	60	60	60	60	30	30	30	30
Dep Var Mean	0.049	0.049	0.049	0.049	32.846	32.846	0.140	0.140
Observations	30946	30946	30946	30946	279	279	279	279

Note: * p<0.1, ** p<0.05, *** p<0.01.
Regressions in columns (1)-(4) are at the household-week level, standard errors clustered by team-week.
Regressions in columns (5)-(8) are at the team-day level (11 days), standard errors clustered by team.
The dependent variable in columns (1)-(4) is a dummy variable indicating whether the visit was conducted alone (see texts for details).
The dependent variable in columns (5)-(6) is the number of sucessful visits conducted by team-day.
The dependent variable in columns (7)-(8) is fraction of visited conducted alone by team-day.
In columns (3)-(4) the selection equation uses the same instrument as in Table 5 (see text for details).

Table 10: Time Use in the Field (Logs)

	(1) Time in Field	(2) Time in Field	(3) Canvassing	(4) Canvassing	(5) Search time	(6) Search time
Horizontal Ethnic Match	-0.035	0.003	0.294**	0.260**	-0.504**	-0.497
	[0.088]	[0.114]	[0.139]	[0.126]	[0.196]	[0.312]
Vertical Ethnic Match		-0.369*		-0.560*		0.053
		[0.192]		[0.274]		[0.348]
R^2	0.067	0.080	0.162	0.181	0.048	0.063
Clusters	30	30	30	30	30	30
Dep Var Mean	5.660	5.660	4.817	4.817	4.873	4.873
Observations	266	266	266	266	266	266

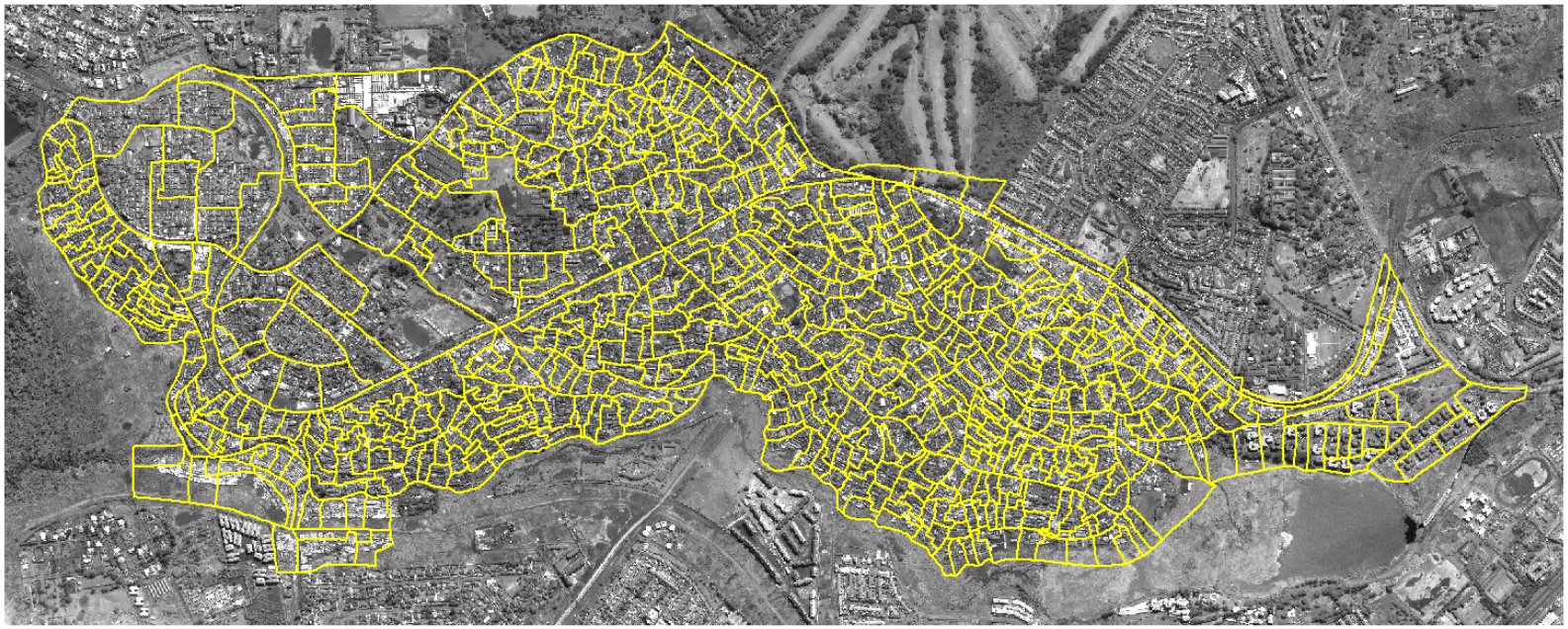
Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team in brackets.
All specifications are team-day level regressions (11 days) and include ethnicity dummies for each staff member.
Idle time is calculated as time spent in field minus canvassing time.
All specifications are in log minutes.

Table 11: Staff Survey Results

	(1) Own Performance	(2) Relative Performance	(3) Complained	(4) Work Hours	(5) Brainstorming
Horizontal Ethnic Match	-0.178	-0.306**	-0.258**	-0.086	5.823
	[0.161]	[0.130]	[0.103]	[0.411]	[9.539]
Vertical Ethnic Match	-0.567***	-0.264	-0.142*	-0.896**	5.067
	[0.194]	[0.239]	[0.083]	[0.418]	[8.237]
Dep var mean	0.492	0.271	0.119	9.127	28.729
Observations	59	59	59	59	59

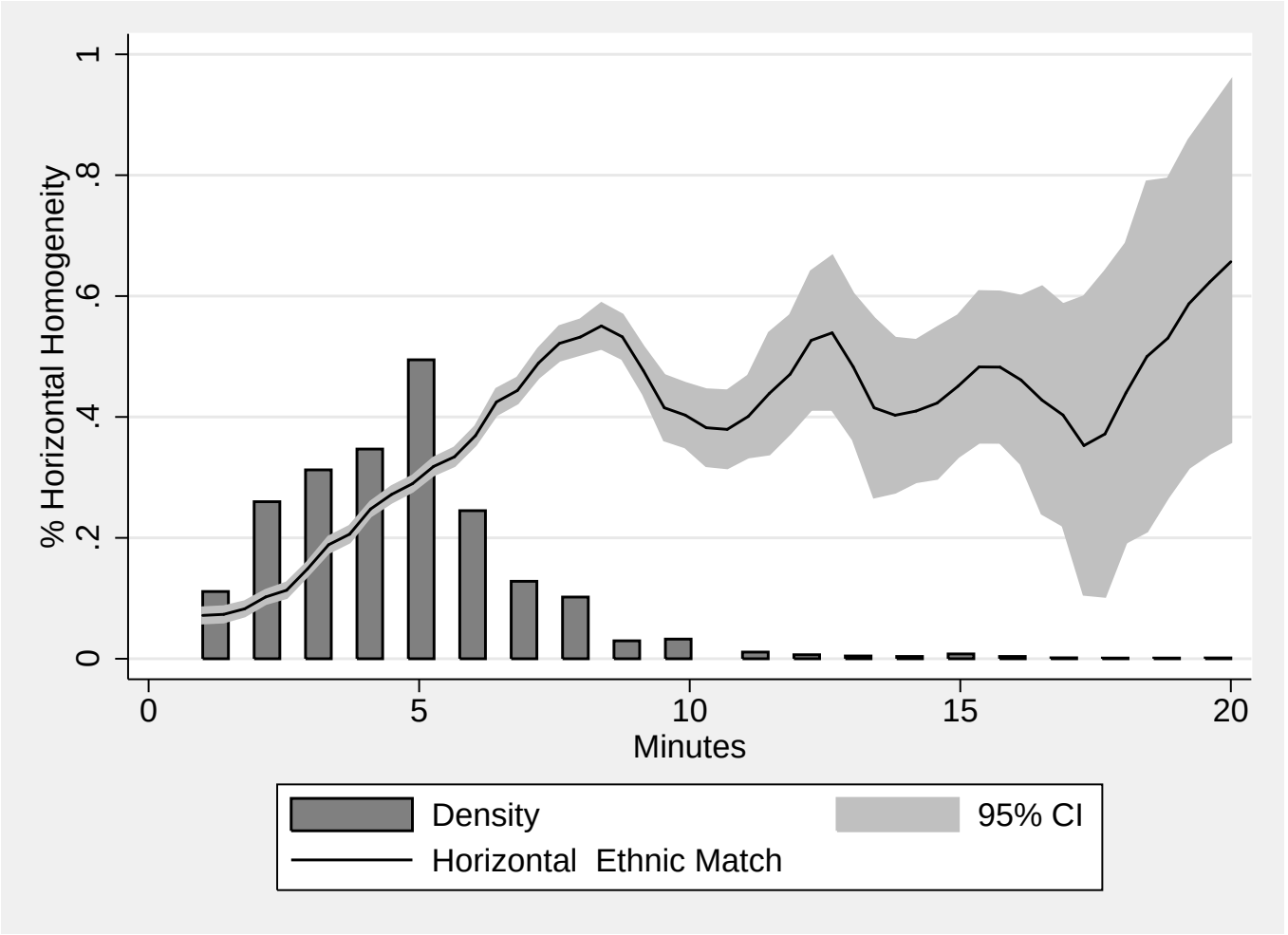
Note: * p<0.1, ** p<0.05, *** p<0.01. Individual-level regressions with robust standard errors in brackets.
Own Performance is a dummy for the canvasser reporting that he/she performed very well during the exercise.
Relative Performance indicates the canvasser reported being the best-performing team member.
Complained indicates the canvasser complained about his/her teammate to his/her manager.
Work hours are self-reported daily work hours.
Brainstorming is the log of the time spent in minutes brainstorming each day.
One staff member could not be contacted during the staff survey.

Figure 1: Map of Kibera



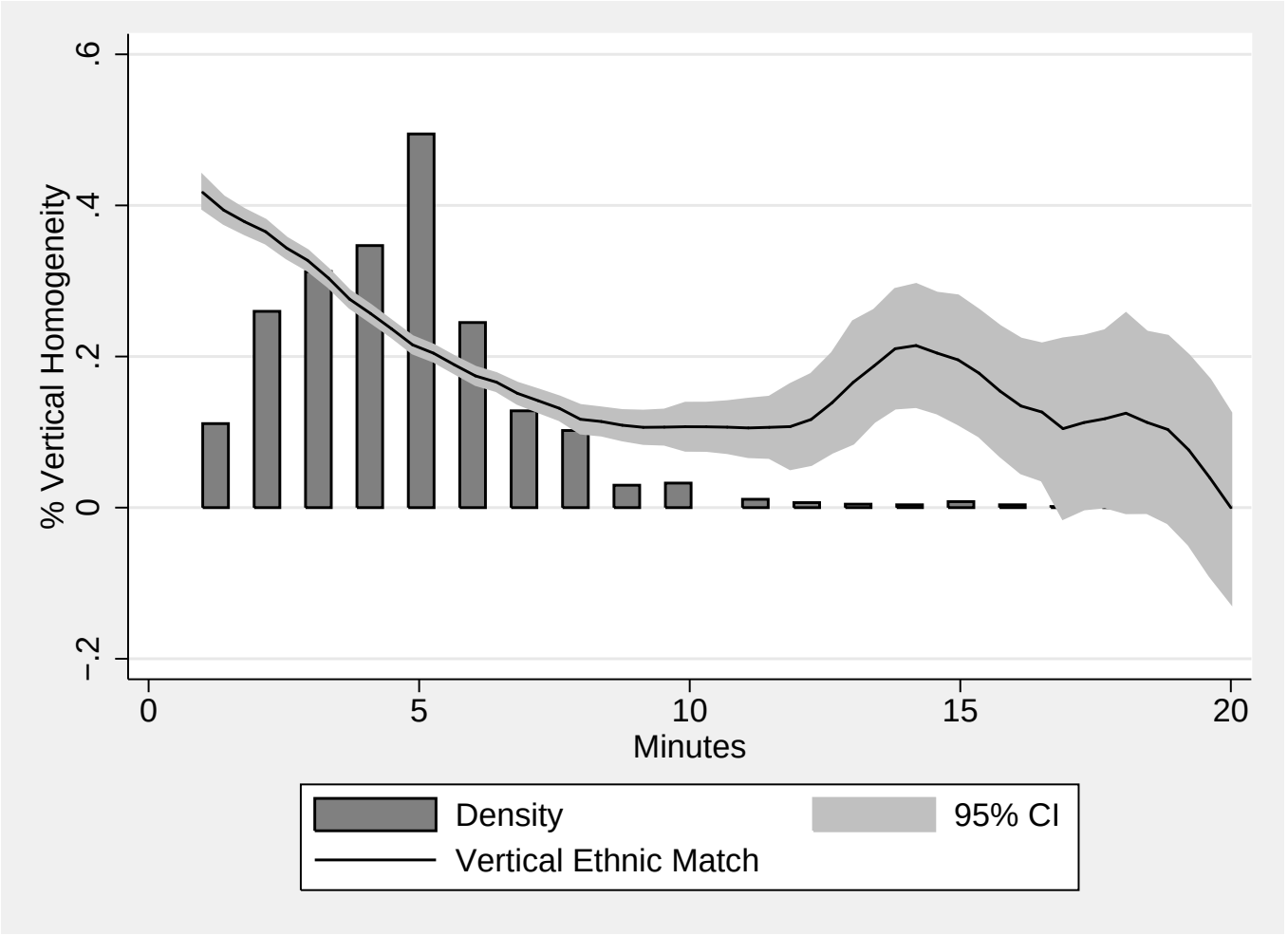
Note: Satellite image of the Kibera slum, with Enumeration Areas (EAs) outlined in yellow.
The Kibera slum in Nairobi covers about five square km of area and over 30,000 households across more than 600 Enumeration Areas.

Figure 2: Horizontal Homogeneity Over Distribution of Duration



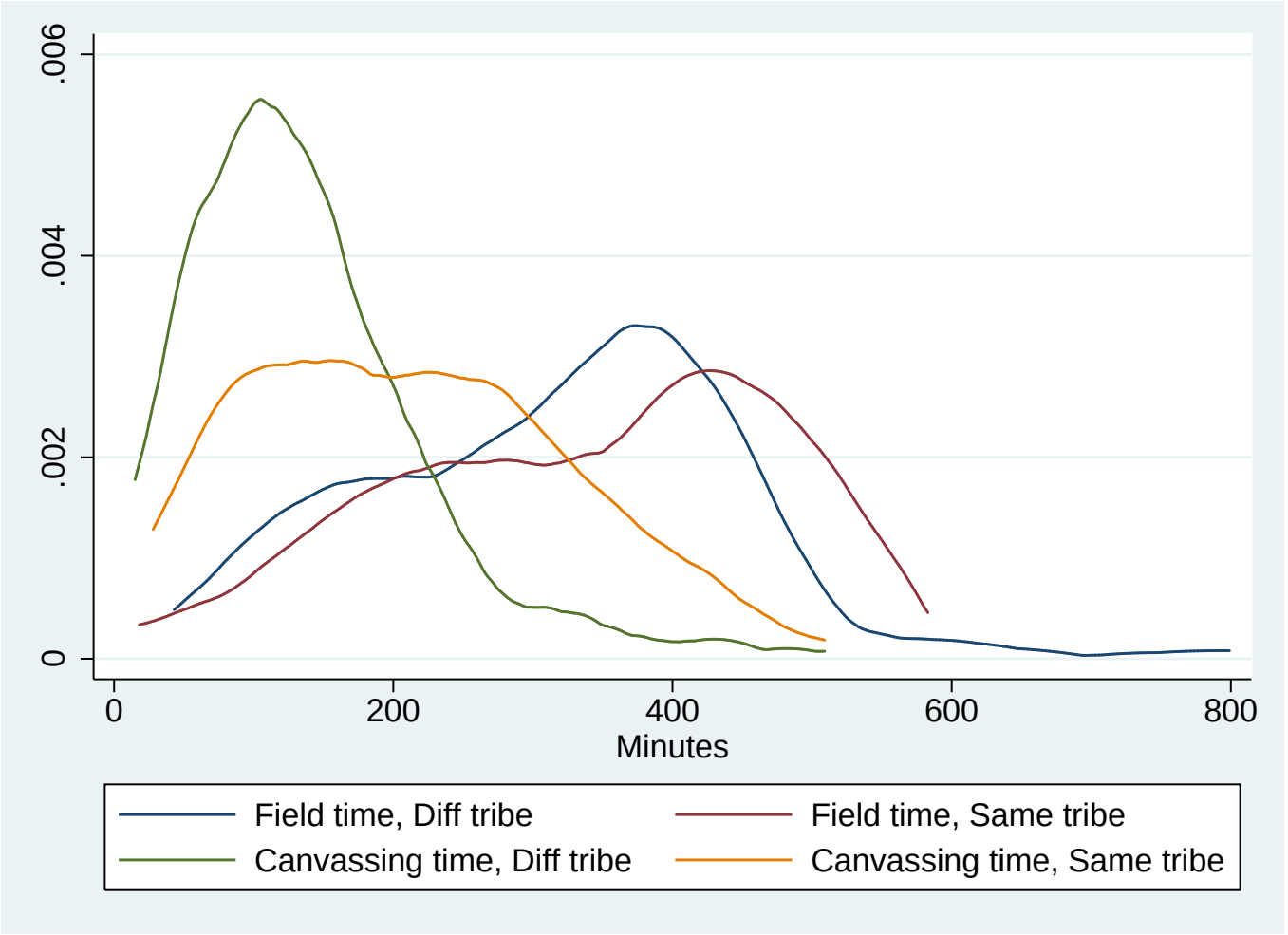
Note: The bar chart is a histogram of visit duration (for completed visits only) for durations 20 minutes or less (99.8% of the data). The solid black line is a kernel density of the horizontal ethnic match over values of visit duration.

Figure 3: Vertical Homogeneity Over Distribution of Duration



Note: The bar chart is a histogram of visit duration (for completed visits only) for durations 20 minutes or less (99.8% of the data). The solid black line is a kernel density of the vertical ethnic match over values of visit duration.

Figure 4: Time Use in the Field, by Horizontal Ethnic Diversity



Note: All plots are kernel densities. Time use is measured in minutes per day.
Canvassing time is measured as the sum of the durations of all visits conducted each day.
Field time is calculated as the difference between the end of the last visit and the beginning of the first visit each day.