Diversity and Team Performance in a Kenyan Organization*

Benjamin Marx
Vincent Pons
Tavneet Suri

This version: March 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 teammate, a 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 effort and performance. The performance effect of ethnic homogeneity within teams is more than twice the size of the effect of homogeneity along gender or age lines. Results from a survey of the canvassers suggest 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 Evan Apfelbaum, Abhijit Banerjee, Nick Bloom, Christina Patterson, Jonathan Petkun, Otis Reid, Michael Stepner, as well as seminar audiences at the MIT Political Economy Lunch, the MIT Sloan Applied Economics Lunch, the Novafrica Seminar, and the University of Capetown. 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 inside the organization - many studies (e.g. Prat (2002), Lazear (1999)) have highlighted this trade-off. 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 try to answer 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 role of three different dimensions of diversity. First, we look at the impact of horizontal diversity (diversity among teammates) on team performance. Second, we look at the effect of vertical (or hierarchical) diversity: we ask whether team performance is enhanced when managers and team members are similar along some dimension. Third, we look at whether similarity between team members and the clients they interact with affects measures of team performance.

We study these questions in the context of a Kenyan non-profit organization that conducts research and collects survey data. The field 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 actual households. Substantial resources are invested in the monitoring of the effort and performance levels of this short-term staff in the field. Making supervision and quality controls effective is therefore crucial to the productivity and long-term success of organizations such as the one we study.

In late 2012, in partnership with this organization, we conducted a voter canvassing exercise around the 2013 Kenyan general election, which was one of a series of experiments that we designed with the Independent Electoral and Boundaries Commission of Kenya (IEBC). The IEBC had a mandate to create a new register of eligible voters across the country before the end of 2012. Our canvassing experiment was designed to encourage citizens to register in one disenfranchised Nairobi neighborhood, called the Kibera slum. The intervention covered approximately 15,000 households in 300 enumeration areas (census tracts)[1] and provided residents with information on the voter 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 staff experiment involved three levels of randomization. First, we randomly allocated the

[1]This is the size of our treatment sample. There was an equal-sized control group where households did not receive any canvassing visit.

canvassers to teams of two. The canvassers were an equal 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 each experienced staff member with a new hire. This layer of randomization induced variation in the degree of horizontal diversity within teams. 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 team members and their managers. Third, each team was randomly allocated a set of enumeration areas (henceforth EAs) to visit in a random order. Fieldwork lasted for two weeks and teams were expected to cover a set of ten EAs each week. This gives us variation in the degree of similarity between canvassers and the households they were supposed to visit. Throughout, we look at three different dimensions of diversity: ethnicity (or tribe),[2] gender, and age.[3]

We measure team performance in two ways. First, we look at whether a particular canvassing visit was completed, i.e. whether the correct household was found and whether the cavasser was able to complete his task. Second, we look at the duration of the 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.

For both measures of performance, we find that teams that are horizontally homogeneous along the ethnicity dimension perform better. This effect is large in magnitude - more than double the size of the effect of team homogeneity along gender or age lines. We find the opposite effect for vertical diversity. Teams that have a manager of a different ethnicity to both teammates perform better, suggesting that even though horizontal team homogeneity may be performance-enhancing, vertical homogeneity is counterproductive in our context. Finally, looking at external diversity, we find that the performance of teams is unaffected by how similar they are to the households on their canvassing list. Given the prevalence of ethnic voting in Kenya, this is a surprising finding as we had expected ethnicity of households to affect 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. We compute the time spent in the field each day, actual time devoted to canvassing households, and the time spent idle or walking between households or dwellings (the last two add up to the first). We interpret time spent in the field as a measure of effort and

[2]We use the words "tribe" and "ethnicity" interchangeably throughout the paper.

[3]We cannot look at measures of diversity in skills or education, as all the staff members had similar levels of education.

the time split between canvassing and non-canvassing as a measure of team productivity. We find that ethnically homogeneous teams spend the same amount of time in the field, more time canvassing, and less time idle. Vertically homogeneous teams spend less time in the field and less time canvassing households, suggesting lower effort levels in those teams.

Second, we conducted a survey of the staff to better understand their practices and behavior during the exercise. We find that teammates who belong to the same ethnicity are less likely to split tasks. The effect is about two thirds of the mean. In our context, splitting tasks meant dividing up the list of households and locations to cover each day across teammates. Splitting was likely not a productive decision, as each team deliberately paired a novice with an experienced staff member who could guide the team around Kibera. This result could partly explain why ethnically homogeneous teams spent less idle time in the field and more time canvassing households. In addition, 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.

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 this literature is beyond the scope of our paper, but a good discussion of the findings outside economics can be found in Apfelbaum, Phillips, and Richeson (2014). In economics, the question of whether and how diversity matters has been studied across fields, in macroeconomic contexts, in microeconomic contexts, in the U.S. and in a number of developing economies. In this section, we highlight some of the studies in these areas and discuss how our paper contributes to the existing literature.

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). Alongside this literature, a number of recent studies focus on the causal impacts of diversity on economic performance, both in firms as well as in other places of economic interactions. Here, we focus on the literature

that deals with diversity in firms and organizations.[4]

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 work with - workers incur disutility from working with people 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, Nickerson, and Owan (2012)). This argument relies on the idea that ethnic diversity brings in diversity of skills - different groups often have access to different (disjoint) information and skill sets, generating complementarities across groups.[5] 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, Nickerson, and Owan (2012)).[6]

Hjort (2014) studies the role of ethnic diversity in team productivity in the Kenyan flower industry. He finds that ethnic conflict increases the negative impact of ethnic diversity on productivity, while the introduction of team pay has the opposite effect. He also argues that these findings must result from preferences à la Becker. His findings, however, are at odds with those of Hoogendoorn, Parker, and van Praag (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.

Lang (1986) argues that the discrimination argument is not borne out in empirical evidence.

[4]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. For example, Fafchamps (2000) and Fisman, Paravisini, and Vig (2012) document the role of ethnic diversity in credit market interactions; Anwar, Bayer, and Hjalmarsson (2012) and Shayo and Zussman (2011) look at legal systems in the U.S. and in Israel, respectively; Fairlie, Hoffmann, and Oreopoulos (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, Baqir, and Easterly (1999), and Miguel and Gugerty (2005)), among others).

[5]Welch (1967) argued that such complementaries existed for example between Blacks and Whites in the US South.

[6]The findings of Hamilton, Nickerson, and Owan (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.

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 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, Nickerson, and Owan (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 co-workers or teammates, rather than the impact of homogeneity across the firm hierarchy. This is at odds with the large literature (e.g., Lazear, Shaw, and Stanton (2014)) highlighting the role managers play in determining team performance. We are able to study both horizontal as well as vertical diversity within the same organization.[7] In addition, we study how external homogeneity, i.e. how similar workers are to the clients of the organization, affects worker performance, again within the same organization.

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 internally cohesive: for example, they are less likely to complain to management about the effort of their teammate, and less likely to report they performed better than their teammate. 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 perfor-

[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 organization, rather than actual hierarchy.

mance 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 3 describes our data. In Section 4, we present the empirical framework and in 5, our main results. We discuss these results in section 6, in light of qualitative evidence from a survey conducted with the canvassing staff. Section 7 concludes.

2 Experimental Design

2.1 Context

A general election was held in Kenya in March 2013. Prior to this election, in the aftermath of a controversial and violent 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 November-December 2012.

To help the IEBC address issues of credibility due to the controversial 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 November-December 2012. This experiment, which we describe in detail below, 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, Collier, and Vicente (2011) and Banerjee, Kumar, Pande, and Su (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, Dormagen, and Pons (2014) for a study in France), we are not aware of any such study in a developing country context, even though registration might represent a significant barrier to electoral participation in these countries.

2.2 Canvassing Experiment

2.2.1 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

therefore 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, as well as the documents 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.

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, Kibera 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, Dormagen, and Pons (2014)). The risk that Kibera voters would be disenfranchised from the electoral process was all the more acute since the area had been a focal point in the 2007 post-election violence. Second, building on earlier fieldwork conducted in the slum (Marx, Stoker, and Suri (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.

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. Figure 1 shows a map of the Kibera area with the main slum subdivisions (called villages) outlined in yellow. 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 being above the median in EA population. Half of the remaining EAs was 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.

2.2.2 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 assistants as well as short-term survey staff. 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, we recruited sixty canvassers for a period of three weeks, including time allocated to training. Thirty of these canvassers had experience and thirty 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, they were carefully trained

by the Principal Investigators 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.

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 monitor 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. The second layer was a team of seven managers, all of whom were previous employees of the organization, though they had worked for the organization for a lot longer than the thirty more experienced canvassers and thus had more advanced fieldwork experience. The sole responsibility of the managers was to spend each day in the field, monitoring and checking on the teams allocated to being under their supervision. They actively watched canvassers in the field and independently checked with households whether the canvassing visit had taken place as planned. Finally, the overall management of the entire field exercise was entrusted to a research manager and a research assistant. The group leaders and managers reported back to the research manager and the research assistant on a daily basis.

To maximize contacts with households, we covered each EA 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. We therefore planned activities so that each household would be visited twice over the two-week period, once each week, and by a different team each time.

2.3 Diversity Experiment

Over the canvassing experiment, we overlaid an experiment on team diversity, with three layers of randomization. The canvassing exercise was conducted by teams of two, in which one experienced canvasser was always paired with one new canvasser. The purpose of this stratification was to build on the experience that the more experienced staff member had acquired in Kibera. As part of the 2012 census (see Section 3), every experienced canvasser had built relationships with local chiefs and elders in the slum, and was therefore expected to conduct fieldwork in Kibera in a safer and more efficient manner. All pairs were therefore randomly composed of one experienced canvasser and one new staff hire, creating exogenous variation in the measure of horizontal diversity within teams. We then randomly allocated four to five teams to a manager, so as to induce random variation in the measure of vertical diversity. 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 they were supposed to canvass. Given the salience of ethnicity in Kenya, especially around elections, we expected that canvassing out-

comes would be different in cases where the team members were of the same ethnicity as the household.

In practice, there were occasional staff replacements due to some canvassers being absent on any given day. In these cases, the group leader would step in to complete the team. Throughout our analysis, we show intent-to-treat estimates based on the initial random composition of teams.

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, a typical team was assigned 532 households over ten EAs during each canvassing week. 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 name of the household head) 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. Since the treatment was allocated at the level of the EA, we wanted the staff to canvass everyone possible in the EA. In our 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, Stoker, and Suri (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.[8] 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 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.

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

3.2 Performance Data

Each canvassing team was provided with tracking sheets that were pre-filled with information on households from our original census data. These tracking sheets included the names of target household heads, along with their GPS coordinates and local area 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.[9]

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. 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 4.7 minutes on average. In our empirical analysis, we use an unconditional measure of time spent with each household, ranging from zero (for unsuccessful visits) to thirty minutes.[10] In addition to the untrimmed version of this variable, we compute a trimmed version where we drop values above the 99th percentile (11 minutes).

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 wages regardless of the number of households successfully canvassed, and the staff was not given specific targets in terms of visit completion rates during training. 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 group leaders monitoring on a daily basis (and sometimes filling in for absent team members), and seven managers in charge of careful monitoring of the progress and the discipline of each team. Third, the measure of duration we use is hard to make up and coordinate on within teams. There was no systematic attempt to give feedback to the staff 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 impossible.

Finally, we use the data from the tracking sheets of each team 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

[9] 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.

[10] We consider as enumeration errors visits recorded to have lasted more than 30 minutes (these represent 0.01 % of the raw tracking data).

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.[11] 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 2012 census, the 2009 census aggregated to the EA level and the tracking sheets. Overall, 41% of all scheduled visits were successfully completed, in line with comparable experiments in the literature (see, for example, Pons and Liegey (2014)).[12] The unconditional visit duration averages 1.79 minutes. 24% of the canvassing teams (7 teams out of 30) were composed of co-ethnic individuals, and 51% were either male-only or female-only teams. 45% of teams had members of similar age (less than two years difference). Lastly, the tribe of manager matches that of one of the team members in 25% of cases. We do not have any instances where the tribe of the manager matches that of the two team members.

3.3 Registration Data

For the original canvassing experiment, we were granted access to 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, and since we did not ourselves collect ID numbers during the initial listing exercise, we are not able to match participants in our experiment to the database of registered voters. To measure treatment effects on voter registration, we therefore rely on self-reported data that we collected from a set of phone calls.

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.[13] 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 as registration outcomes.

3.4 Staff Survey Data

We conducted a survey of the canvassing staff a year after the completion of the experiment. We use data from this survey as qualitative evidence to support our main experimental results. The survey questionnaire was conducted individually and collected data on social interactions

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

[12] The average success rate was 43.5% for co-ethnic teams and 40.7% for non-coethnic teams.

[13] 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.

between teammates (e.g., did teammates interact socially after work or after the experiment), working methods (e.g., did the team split tasks, or what language was primarily used within the team), as well as self-reported measures of working hours and their performance relative to their teammate. One of the canvassers could not be located at the time of this survey, bringing the number of respondents to fifty-nine.

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 largely focus on the impact of ethnic homogeneity (horizontal and vertical), though we also briefly discuss results for the impact of gender and age homogeneity for comparison.

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 homogeneity in the team along ethnicity lines, 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 tribe 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.

Note that we only use households allocated to 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 these EAs.

In each table, 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, 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.

We compare the magnitude of the effect of horizontal ethnic homogeneity with the effect of homogeneity by gender or age. For homogeneity by gender, the specification we use is:

$$y_{ijt} = \alpha + \beta GM_{jt} + \delta^k_{jt} + \delta^l_{jt} + \Omega X_{ij} + \epsilon_{ijt} \quad (2)$$

where GM_{jt} is a dummy equal to one if the team has either two males or two females; δ^k_{jt} and δ^l_{jt} are two sets of gender dummies, one for each teammate, and other variables are defined as above. For homogeneity by age, we look at a similar specification:

$$y_{ijt} = \alpha + \beta AM_{jt} + \zeta^k_{jt} + \zeta^l_{jt} + \Omega X_{ij} + \epsilon_{ijt} \quad (3)$$

where AM_{jt} is a dummy equal to one if the ages of the two teammates are within two years of each other; ζ^k_{jt} and ζ^l_{jt} are the ages of each teammate.

In addition to the tests of horizontal homogeneity, we also look at whether vertical homogeneity matters, i.e. whether the ethnicity of the manager matters. There are no cases in the experiment where the tribe of the two team members and that of the manager all match. However, in 25% of cases the manager belongs to the same tribe as one of the team members. We look at whether this affects team performance, using the following specification:

$$y_{ijt} = \alpha + \beta_1 EM_{jt} + \beta_2 MEM_{jt} + \gamma^k_{jt} + \gamma^l_{jt} + \gamma^m_{jt} + \Omega X_{ij} + \epsilon_{ijt} \quad (4)$$

where MEM_{jt} is a dummy variable equal to one if the ethnicity of the manager matches the ethnicity of either team member; γ^k_{jt}, γ^l_{jt} and γ^m_{jt} are three sets of tribe dummies, one for each of the two team members and one for the manager. The correlation between the EM_{jt} and the MEM_{jt} dummies is -0.32 in our data.

Finally, we look at whether external homogeneity matters, i.e. whether the ethnicity of households matters for the performance of the canvassers. For this, we use the following specification:

$$y_{ijt} = \alpha + \beta_1 EM_{jt} + \beta_2 HM_{ijt} + \beta_3 HMM_{ijt} + \gamma^k_{jt} + \gamma^l_{jt} + \gamma^m_{jt} + \gamma^n_i + \Omega X_{ij} + \epsilon_{ijt} \quad (5)$$

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^n_i are ethnicity dummies for household i.

Note there are at least two 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 (ethnic voting is particularly salient in Kenya). 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 of the 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 from regressing each relevant household and EA level characteristic on the horizontal homogeneity team dummy (which equals one for team members of the same ethnicity). For EA-level variables, we run EA-level regressions. 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 one of the team members) as the main regressor. In Column (4), we report similar results for horizontal homogeneity along gender lines, and in column (5) for horizontal homogeneity along age lines.

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 a dummy for households with a phone number, household size, the (log) amount paid in housing rent, the number of years spent in the same structure, and the number of years spent in Kibera. The latter include the age, gender and education of the household head, as well as various measures of occupation, employment, and socio-economic status (consumption and poverty).[14] The regressions using variables collected in Marx, Stoker, and Suri (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. Only one coefficient (out of 15) is significant in each one of columns (2), (3) and (5). Column (4) - where the same-gender dummy is the main regressor - looks a little worse since four variables are significant at conventional levels.

5.2 Main Results

The estimates from equation (1) are presented in Tables 3a and 3b for the two outcomes of interest: 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 impact of a team's horizontal ethnic homogeneity on performance. We show estimates obtained from different specifications in Table 3a,

[14] The poverty indicator was computed as part of a poverty mapping exercise. See Marx, Stoker, and Suri (2014) for details.

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 45% 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 (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 we have a number of 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, where 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, given the small number of clusters. Throughout, horizontal ethnic homogeneity has a strong positive effect on performance.

5.3 Homogeneity by Gender and Age

In Table 4, we look at the performance effects of team homogeneity by gender and age. Standard errors are reported for two alternative types of clustering: clustering by team-week and clustering by EA. We do find that gender-homogeneous teams appear to perform better (columns (1)-(4)). This is true for both outcomes of interest, the completion of canvassing visits and the duration of the visits, though the magnitudes are smaller. 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 7% longer with a gender match. The coefficient on visit duration

is not statistically different from zero when we cluster by team-week, our preferred clustering method. 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 tenth the size for duration.

Columns (5)-(8) of Table 4 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.

5.4 Manager Matches

Given the fact that managers were randomly assigned to teams, we also exploit exogenous variation in whether there was vertical diversity, i.e. whether the manager and team members 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. Instead, we look at whether the ethnicity of the manager matches the ethnicity of any one of the team members. Table 5 reports these results. Looking at both measures of team performance, we find evidence of effects that are opposite in sign to the effect of horizontal ethnic homogeneity. Remembering that this effect is identified from teams that are not homogeneous by tribe themselves, the magnitudes 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 using the wild bootstrap.

5.5 Ethnic matches with households

In Table 6, we look at the impact of external diversity, i.e. whether the households' ethnicity affects the success and duration of canvassing visits (equation (5)). The effect of an external ethnic match is a precisely estimated zero, in that not only is it not statistically significant but we can also rule out 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 happens only in 3% of cases. This effect is negative, small in magnitude, and not significantly different from zero (results available upon request).

5.6 Effects on Registration

In Table 7, we look at the longer-term impacts of the horizontal ethnic diversity of canvassing teams. We use as our dependent variable the self-reported measure of household registration

described above. 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 80%).

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 effect is 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 the administrative registration data suggests the canvassing intervention 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 effort and performance, and the second on outcomes collected from our survey of the canvassers.

In Table 8 and Figure 2, we present results for effort and performance. 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 idle or walking between EAs, which is simply the difference between the first two measures. As mentioned above, 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. Figure 2 shows the distributions 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 8 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 idle or walking between locations. Teams that have a vertical ethnic match with the manager spend both less time in the field and less time canvassing. This is suggestive evidence that effort levels are lower within those teams.

In Table 9, we present results from the survey conducted with the canvassing staff. These surveys 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, and present results in Table 9. As with earlier specifications, we control for tribe main effects.

In Column (1) of Table 9, we show that teams that matched horizontally on ethnicity were significantly less likely to split the team each day. The split variable comes from a question in the survey that asked how often they split the team to conduct interviews individually (ranging from every day to never). The dependent variable is a dummy for whether the canvasser reported that they split the team either every day all the time or every day part of the day (as opposed to only on some days, very rarely or never). During training we did not give clear instructions as to whether splitting was discouraged or not. However, within each team we deliberately paired a more experienced staff member with a new staff member so as to promote synergies and learning. In addition, splitting meant the novice canvasser would not benefit from the more experienced staff member guiding the team around the slum, and that the two teammates could not observe and monitor each other's effort and performance in the field. Splitting was therefore likely to be harmful to performance.

In columns (2) and (3) of Table 9, 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 (2), 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, where we asked canvassers to report whether they performed better than their teammate or vice versa. In column (3), we report results for this variable. Canvassers that were part of a team that was horizontally homogeneous along ethnicity lines 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 (4), we show that horizontally matched teams were significantly less likely to complain about their teammate, as were vertically matched teams. In columns (5) and (6) we look at self-reported measures of effort. Hours worked are no different in horizontally homogeneous teams (consistent with the results in Table 8), but are significantly lower for vertically matched teams. In column (6), we look at the time the team spent brainstorming. We find no differences across homogeneous and diverse teams along this metric.

Finally, we should note that preferences towards working with similar workers are unlikely to explain our results for two reasons. First, we do not find that canvassers are more likely to find or more likely to spend more time with households of the same ethnicity. Second, in the canvasser survey, we find that teammates of the same ethnicity were not 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 this, while all others communicated in either English or Swahili, the *lingua franca* of Kenya.

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 and hence performance. Meanwhile, the vertical homogeneity led to less stringent norms and discipline and hence lower effort and performance. These results also cement the importance of looking at different dimensions of diversity - there may be costs to diversity in terms of the horizontal structure of the firm. But this is balanced against the large gains to diversity we find along the vertical structure of the organization.

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 ahead of a major election in 2013. 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, 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 teammate and the manager (the vertical dimension) has the opposite effect: teams that share their ethnicity with the manager perform worse. 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 canvassing staff. First, the survey helps us rule out simple preference-based or language expla-

nations 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 less likely to split the team. Canvassers in vertically homogeneous teams report working (significantly) less hours and provide a more negative assessment of their own performance.

Our findings suggest much 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 may induce more effort and/or different norms within teams.

References

AKER, J. C., P. COLLIER, AND P. C. VICENTE (2011): "Is Information Power? Using Cell Phones During an Election in Mozambique," *Working Paper*.

ALESINA, A., R. BAQIR, AND W. EASTERLY (1999): "Public Goods and Ethnic Divisions," *Quarterly Journal of Economics*, 114, 1243–1284.

ALESINA, R., AND E. L. FERRARA (2005): "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, 43(3), 762–800.

ANWAR, S., P. BAYER, AND R. HJALMARSSON (2012): "The Impact of Jury Race in Criminal Trials," *Quarterly Journal of Economics*, 127(2), 1017–1055.

APFELBAUM, E. P., K. W. PHILLIPS, AND J. A. RICHESON (2014): "Rethinking the Baseline in Diversity Research: Should We Be Explaining the Effects of Homogeneity," *Perspectives on Psychological Science*, 9(3), 235–244.

ARROW, K. (1973): "The Theory of Discrimination," in *Discrimination in Labor Markets*, ed. by O. Ashenfelter, and A. Rees. Princeton University Press, Princeton, NJ.

BANERJEE, A. V., S. KUMAR, R. PANDE, AND F. SU (2010): "Do Informed Voters Make Better Choices? Experimental Evidence from Urban India," *Working Paper*.

BECKER, G. (1957): *The Economics of Discrimination*. University of Chicago Press.

BRACONNIER, C., J.-Y. DORMAGEN, AND V. PONS (2014): "Voter Registration Costs and Disenfranchisement: Experimental Evidence from France," *Working Paper*.

EASTERLY, W., AND R. LEVINE (1997): "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics*, 112(4), 1203–1250.

FAFCHAMPS, M. (2000): "Ethnicity and Credit in African Manufacturing," *Journal of Development Economics*, 61, 205–235.

FAIRLIE, R. W., F. HOFFMANN, AND P. OREOPOULOS (2014): "A Community College Instructor Like Me: Race and Ethnicity Interactions in the Classroom," *American Economic Review*, 104(8), 2567–2591.

FISMAN, R., D. PARAVISINI, AND V. VIG (2012): "Cultural Proximity and Loan Outcomes," *NBER Working Paper No. 18096*.

HAMILTON, B. H., J. A. NICKERSON, AND H. OWAN (2012): "Diversity and Productivity in Production Teams," *Advances in the Economic Analysis of Participatory and Labor-Managed Firms*, 13, 99–138.

HJORT, J. (2014): "Ethnic Divisions and Production in Firms," *Quarterly Journal of Economics*, 129(4).

HOOGENDOORN, S., S. C. PARKER, AND M. VAN PRAAG (2012): "Ability Dispersion and Team Performance: A Field Experiment," *Tinbergen Institute Discussion Paper 2012-130/VII*.

HOOGENDOORN, S., AND M. VAN PRAAG (2014): "Ethnic Diversity and Team Performance: A Field Experiment," *Tinbergen Institute Discussion Paper 2012-068/3*.

KANDEL, E., AND E. P. LAZEAR (1992): "Peer Pressure and Partnerships," *Journal of Political Economy*, 100(4), 801–817.

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

LAZEAR, E. P. (1999): "Globalisation and the Market for Team-Mates," *The Economic Journal*, 109, C15–C40.

LAZEAR, E. P., K. L. SHAW, AND C. T. STANTON (2014): "The Value of Bosses," *forthcoming Journal of Labor Economics*.

MARX, B., T. STOKER, AND T. SURI (2014): "There Is No Free House: Ethnic Patronage and Diversity in a Kenyan Slum," *Working Paper*.

MAS, A., AND E. MORETTI (2009): "Peers at Work," *American Economic Review*, 99(1), 112–145.

MIGUEL, E., AND M. K. GUGERTY (2005): "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya," *Journal of Public Economics*, 89, 2325–2368.

PONS, V., AND G. LIEGEY (2014): "Increasing the Electoral Participation of Immigrants: Experimental Evidence from France," *Working Paper*.

PRAT, A. (2002): "Should a Team be Homogeneous?," *European Economic Review*, 46, 1187–1207.

PRICE, J., AND J. WOLFERS (2010): "Racial Discrimination among NBA Referees," *The Quarterly Journal of Economics*, 125(4), 1859–1887.

REINHARD, S., AND M. WARGLIEN (2007): "The Emergence of Simple Languages in an Experimental Coordinate Game," *PNAS*, 104(18), 7361–7366.

SHAYO, M., AND A. ZUSSMAN (2011): "Judicial Ingroup Bias in the Shadow of Terrorism," *Quarterly Journal of Economics*, 126, 1447–1484.

WELCH, F. (1967): "Labor Market Discrimination: An Interpretation of Income Differences in the Rural South," *Journal of Political Economy*, 75, 225–240.

ZENGER, T. R., AND B. S. LAWRENCE (1989): "Organizational Demography: The Differential Effects of Age and Tenure Distributions on Technical Communication," *Academy of Management Journal*, 32(2), 353–376.

Table 1: Summary Statistics

	Mean	SD	*N*
Household Characteristics (Authors' Census of Kibera Slum)			
Household Has Phone	0.51	0.50	31352
Household Size	3.67	2.13	31302
Years in Same Structure	8.23	8.71	16916
Years in Kibera	15.55	10.95	16818
EA Characteristics (National Bureau of Statistics Census)			
Number of Households in EA	52.25	26.25	300
Average Age of Household Head	35.40	2.19	298
Gender of Household Head (1=Male)	0.81	0.08	298
Household Head has No Education	0.04	0.05	298
Household Head only has Primary Education	0.47	0.14	298
Household Head Unemployed	0.32	0.16	298
Household Head Owns a Business	0.15	0.12	298
Log Consumption per Capita	10.42	0.20	298
EA Poverty Rate	0.15	0.07	298
Canvassing Visit Level Data			
Visit Completed	0.41	0.49	30936
Any Additional Individuals Canvassed	0.01	0.11	31352
Duration of Visit (minutes)	1.79	2.83	29381
Duration of Visit, Trimmed (minutes)	1.67	2.49	29149
Duration, Conditional on Finding Household	4.73	2.69	11116
Daily Time Spent Canvassing (Minutes)	194.6	143.3	31312
Daily Time Spent in Field (Minutes)	345.8	125.2	23332
Horizontal Ethnic Match	0.24	0.42	31352
Vertical Ethnic Match	0.25	0.43	31352
Horizontal Gender Match	0.51	0.50	31352
Horizontal Age Match	0.45	0.50	31352
Total Distance Assigned to Team	10.83	1.79	31352
Total Sample Assigned to Team	532.4	69.7	31352
External Ethnic Match: Canvasser	0.27	0.44	31352
External Ethnic Match: Manager	0.16	0.36	31352
Canvassing Team Level Data			
Total Households Allocated to Team	1045.1	145.9	30
Total Households Visited by Team	425.6	64.5	30
Total EAs Allocated to Team	20.00	0.53	30

Note: The author's census of Kibera was conducted in early 2012.

The EA characteristics are aggregated from the micro data for the 2009 national census.

Each team was composed of two staff members. There were 30 teams in total.

Each team covered 10 EAs in each week for a total of two weeks (and 20 total EAs).

Each household was visited twice, each time by a different team.

Table 2: Randomization Checks

	Mean	Horizontal Ethnic Match	Vertical Ethnic Match	Horizontal Gender Match	Horizontal Age Match
Household Has Phone	.5125	.0299	.0257	.0029	-.0344**
		[.0195]	[.0375]	[.0126]	[.0148]
Household Size	3.6692	.0029	.0542	-.0439	-.0638
		[.0538]	[.0888]	[.0393]	[.0421]
Housing Rent (log)	7.2695	-.0498	.0699	-.0383	.0148
		[.0553]	[.0877]	[.0375]	[.034]
Years in Same Structure	8.2269	.196	.1559	.2528	-.0203
		[.2543]	[.8406]	[.1892]	[.2238]
Years in Kibera	15.5485	.1173	-.087	.2822	-.0487
		[.3555]	[1.1647]	[.2389]	[.3125]
Number of Households in EA	52.2533	.9767	1.6603	-4.857**	-3.6712
		[2.7272]	[7.5348]	[1.8793]	[2.2617]
Average Age of Household Head	35.4021	-.0496	-.0245	.0928	-.0462
		[.2741]	[.7055]	[.1689]	[.2133]
Gender of Household Head	.814	.0026	-.0048	.0098*	.0077
		[.0087]	[.0165]	[.0056]	[.006]
Household Head has No Education	.0352	.0048	-.0217***	.0053	.0034
		[.0053]	[.0082]	[.0034]	[.0041]
Household Head has Primary Education	.4711	-.031**	.0104	-.0047	-.0047
		[.0126]	[.0306]	[.01]	[.0108]
Household Head Owns a Business	.1496	-.016	-.0122	.0138	.0002
		[.0134]	[.0338]	[.0094]	[.0108]
Household Head in Private Sector	.4235	-.0039	-.0236	.0361**	.0035
		[.021]	[.0562]	[.0153]	[.0193]
Household Head Unemployed	.3157	-.0194	-.0103	.0207**	.003
		[.0161]	[.0349]	[.01]	[.0102]
Log Consumption per Capita	10.4168	.0061	.0456	-.0062	.0029
		[.0215]	[.0505]	[.0154]	[.0171]
EA Poverty Rate	.1467	-.0053	-.0273	-.0044	-.0034
		[.0096]	[.0248]	[.0062]	[.0073]

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team-week in brackets.
In columns (2)-(5), each cell shows estimates from a regression of the row variable on the column variable and relevant tribe dummies.
Column (1) shows the mean of each row variable.

Table 3a: Ethnic Composition and Team Performance

	(1) Visit Completed	(2) Visit Completed	(3) Visit Duration	(4) Visit Duration	(5) Visit Duration (trimmed)
Horizontal Ethnic Match	0.066***	0.091***	0.813***	0.812***	0.729***
	[0.023]	[0.027]	[0.212]	[0.223]	[0.211]
Random Visit Order		-0.004***		-0.034***	-0.032***
		[0.001]		[0.009]	[0.008]
IEBC Treatment		-0.048**		0.012	-0.046
		[0.020]		[0.161]	[0.153]
EA population		-0.008		-0.039	-0.057
		[0.009]		[0.051]	[0.047]
R^2	0.008	0.011	0.025	0.030	0.032
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	1.788	1.788	1.000
Observations	30936	30936	29381	29381	29149

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 11 minutes).

Table 3b: 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	0.091***	0.091***	0.091***	0.091***	0.091***	0.812***	0.812***	0.812***	0.812***	0.812***
	[0.027]	[0.030]	[0.016]	[0.026]	[0.028]	[0.223]	[0.282]	[0.108]	[0.215]	[0.273]
Random Visit Order	-0.004***	-0.004***	-0.004***	-0.004***	-0.004***	-0.034***	-0.034***	-0.034***	-0.034***	-0.034***
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.009]	[0.007]	[0.006]	[0.009]	[0.007]
IEBC Treatment	-0.048**	-0.048**	-0.048***	-0.048**	-0.048**	0.012	0.012	0.012	0.012	0.012
	[0.020]	[0.020]	[0.015]	[0.021]	[0.020]	[0.161]	[0.200]	[0.081]	[0.156]	[0.193]
EA population	-0.008	-0.008	-0.008	-0.008	-0.008	-0.039	-0.039	-0.039	-0.039	-0.039
	[0.009]	[0.009]	[0.012]	[0.010]	[0.010]	[0.051]	[0.055]	[0.062]	[0.050]	[0.053]
R^2	0.011	0.011	0.011	0.011	0.011	0.030	0.030	0.030	0.030	0.030
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		.006					.022			
Dep Var Mean	0.413	0.413	0.413	0.413	0.413	1.788	1.788	1.788	1.788	1.788
Observations	30936	30936	30936	30936	30936	29381	29381	29381	29381	29381

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: Gender and Age Composition and Team Performance

	(1) Completed	(2) Completed	(3) Duration	(4) Duration	(5) Completed	(6) Completed	(7) Duration	(8) Duration
Horizontal Gender Match	0.037**	0.037***	0.123	0.123*				
	[0.017]	[0.012]	[0.139]	[0.068]				
Horizontal Age Match					0.035*	0.035**	-0.052	-0.052
					[0.019]	[0.014]	[0.183]	[0.084]
R^2	0.005	0.005	0.011	0.011	0.008	0.008	0.012	0.012
Clustering by Team-Week	X		X		X		X	
Clustering by EA		X		X		X		X
Clusters	60	300	60	300	58	300	58	300
Dep Var Mean	0.413	0.413	1.788	1.788	0.414	0.414	1.780	1.780
Observations	30936	30936	29381	29381	30146	30146	28627	28627

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

Table 5: Ethnic Composition and Team Performance: Management

	(1) Completed	(2) Completed	(3) Completed	(4) Duration	(5) Duration	(6) Duration
Vertical Ethnic Match	-0.058	-0.058	-0.058	-1.339***	-1.339***	-1.339***
	[0.038]	[0.045]	[0.039]	[0.283]	[0.351]	[0.210]
Horizontal Ethnic Match	0.079***	0.079***	0.079***	0.758***	0.758***	0.758***
	[0.029]	[0.022]	[0.021]	[0.213]	[0.235]	[0.128]
R^2	0.013	0.013	0.013	0.033	0.033	0.033
Clustering by Team-Week	X			X		
Clustering by Team		X			X	
Clustering by EA			X			X
Clusters	60	30	300	60	30	300
Wild Bootstrap P-value: Vertical		.380			.002	
Wild Bootstrap P-value: Horizontal		.064			.044	
Dep Var Mean	0.413	0.413	0.413	1.788	1.788	1.788
Observations	30936	30936	30936	29381	29381	29381

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

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

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

Table 6: 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	0.006	0.006	0.006
	[0.008]	[0.008]	[0.009]	[0.049]	[0.050]	[0.055]
External Ethnic Match: Manager	0.003	0.003	0.003	-0.091	-0.091	-0.091
	[0.009]	[0.010]	[0.011]	[0.064]	[0.070]	[0.069]
Horizontal Ethnic Match	0.083***	0.083**	0.083***	0.778***	0.778***	0.778***
	[0.021]	[0.034]	[0.020]	[0.260]	[0.241]	[0.127]
R^2	0.022	0.022	0.022	0.035	0.035	0.035
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	0.423	0.423	0.423	1.838	1.838	1.838
Observations	28223	28223	28223	26790	26790	26790

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 2,752 households in our sample.

Table 7: 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**	0.872***	-0.016	-0.005	0.008	-0.007
	[0.033]	[0.256]	[0.012]	[0.017]	[0.009]	[0.017]
R^2	0.008	0.035	0.002	0.002	0.004	0.006
Clusters	60	60	60	60	60	60
Dep var mean	0.488	2.131	0.799	0.686	0.050	0.585
Observations	6717	6334	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 8: Time Use in the Field

	(1) Time in Field	(2) Time in Field	(3) Canvassing	(4) Canvassing	(5) Idle time	(6) Idle time
Horizontal Ethnic Match	-0.039	-0.011	0.276*	0.241	-0.456***	-0.409*
	[0.086]	[0.110]	[0.153]	[0.144]	[0.147]	[0.225]
Vertical Ethnic Match		-0.359*		-0.601*		-0.154
		[0.188]		[0.319]		[0.280]
R^2	0.069	0.081	0.158	0.181	0.062	0.079
Clusters	30	30	30	30	30	30
Dep Var Mean	5.649	5.649	4.780	4.780	4.933	4.933
Observations	267	267	267	267	265	265

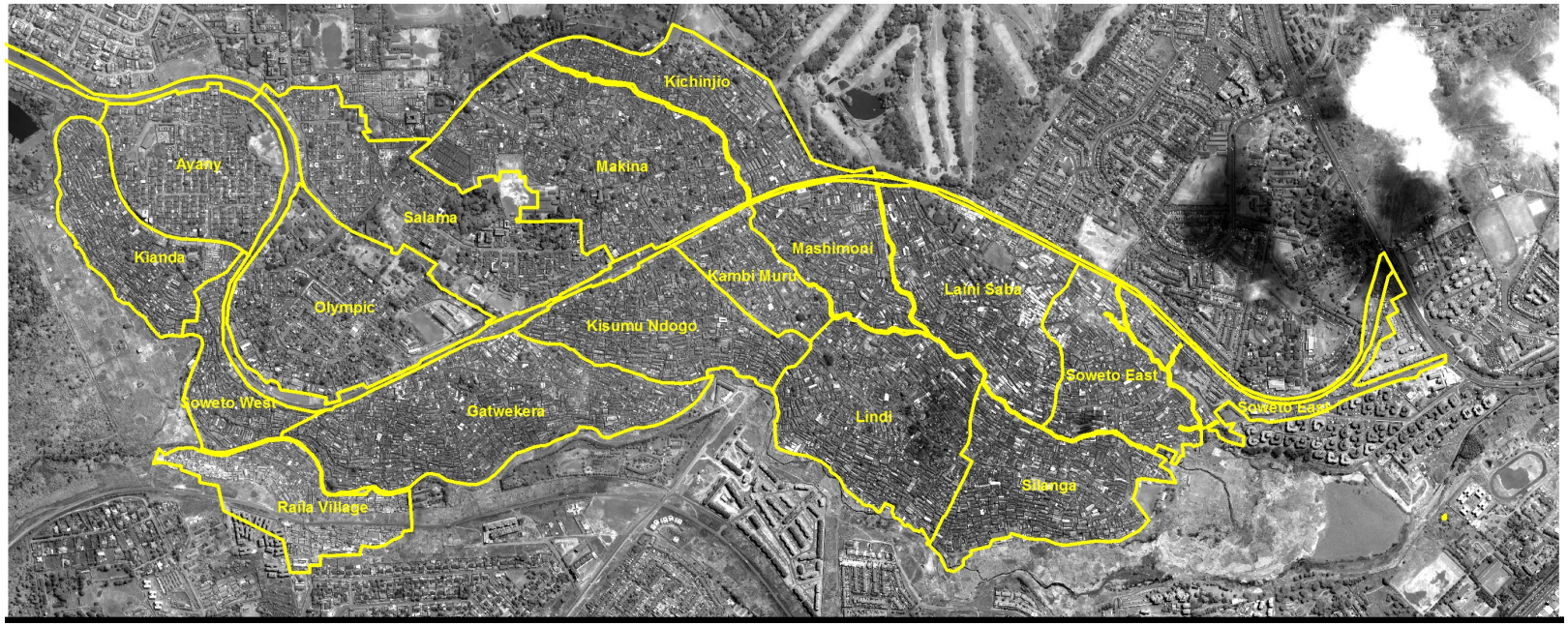
Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team in brackets.
All specifications are team-day level regressions (11 days) and include ethnicity dummies for each staff member.
Idle time is calculated as time spent in field minus canvassing time.
All specifications are in log minutes.

Table 9: Staff Survey Results

	(1) Split Team	(2) Own Performance	(3) Relative Performance	(4) Complained	(5) Work Hours	(6) Brainstorming
Horizontal Ethnic Match	-0.409**	-0.178	-0.306**	-0.258**	-0.086	0.071
	[0.177]	[0.161]	[0.130]	[0.103]	[0.411]	[0.308]
Vertical Ethnic Match	-0.169	-0.567***	-0.264	-0.142*	-0.896**	0.191
	[0.219]	[0.194]	[0.239]	[0.083]	[0.418]	[0.328]
Dep var mean	0.644	0.492	0.271	0.119	9.127	3.218
Observations	59	59	59	59	59	58

Note: * p<0.1, ** p<0.05, *** p<0.01. Individual-level regressions with robust standard errors in brackets.
Split Team indicates canvassers split the team at least part of every day to conduct interviews individually.
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.

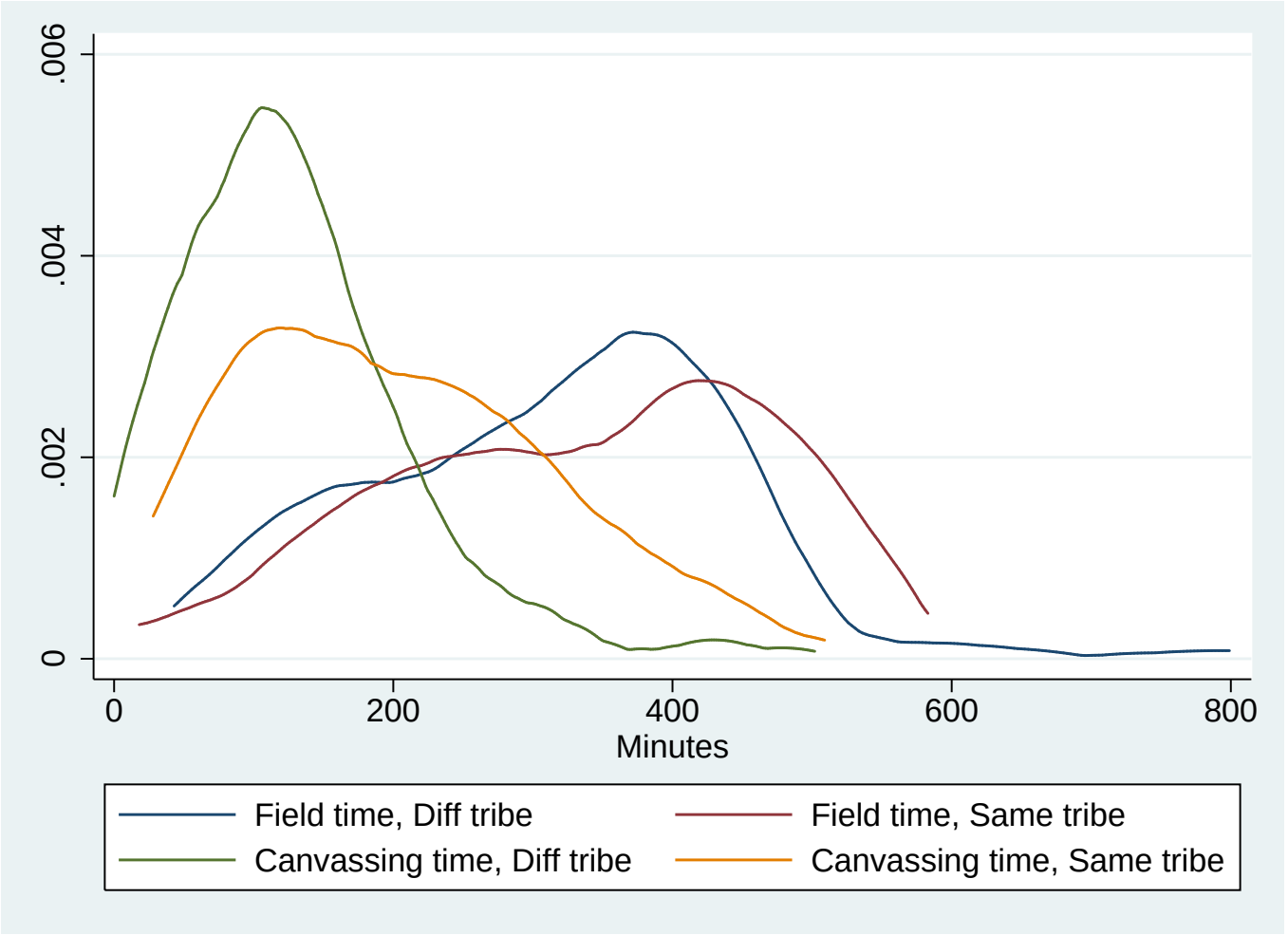
Figure 1: Map of Kibera



Note: Satellite image of the Kibera slum, with villages in Kibera 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: 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.