Why the Referential Treatment? Evidence from Field Experiments on Referrals

Amanda Pallais

Harvard University and National Bureau of Economic Research

Emily Glassberg Sands

Coursera

Referred workers are more likely than nonreferred workers to be hired, all else equal. In three field experiments in an online labor market, we examine why. We find that referrals contain positive information about worker performance and persistence that is not contained in workers' observable characteristics. We also find that referrals perform particularly well when working directly with their referrers. However, we do not find evidence that referrals exert more effort because they believe their performance will affect their relationship with their referrer or their referrer's position at the firm.

I. Introduction

A large empirical literature has shown that the majority of jobs are found through informal contacts, firms are more likely to hire applicants re-

We would like to thank David Autor, Felipe Barrera-Osorio, Patrick Bayer, Raj Chetty, Melissa Dell, David Deming, Itzik Fadlon, Adam Guren, John Friedman, Roland Fryer, Edward Glaeser, Claudia Goldin, Josh Goodman, Rick Hornbeck, Lisa Kahn, Lawrence Katz, John List, Ben Schoefer, Sarah Turner, Marty West, seminar participants at Berkeley, Booth, Brookings Institution, Columbia, Duke, Harvard, Kellogg, and the National Bureau of Economic Research Summer Institute Labor Studies, the New York Federal Reserve, Princeton, RAND, University of British Columbia, University of Chicago, and Wharton, as well as Jesse Shapiro and four anonymous referees for their many helpful comments and suggestions. We would like to thank John Horton and the oDesk Corporation for help running the ex-

Electronically published October 31, 2016
[*Journal of Political Economy*, 2016, vol. 124, no. 6]

ferred by current employees than nonreferred applicants, and some firms even give bonuses to employees for successful referrals (see, e.g., Granovetter 1995; Fernandez and Weinberg 1997; Bewley 1999; Peterson, Saporta, and Seidel 2000; Ioannides and Loury 2004; Topa 2011; Brown, Setren, and Topa 2012; Burks et al. 2015). Yet the literature remains divided on why firms draw so heavily on referred applicants. Referrals may provide (positive) information about worker quality, or being referred may induce a worker to work harder or more productively; alternatively, firms may hire referrals for nepotistic reasons or to decrease recruiting costs (see, e.g., Montgomery 1991; Simon and Warner 1992; Kugler 2003; Wang 2013; Heath 2015). This paper analyzes a set of field experiments in an online labor market to answer two open questions about referrals: First, do referrals contain information about worker productivity? Second, do referred workers work harder or more effectively because they are referred?

Answering the first of these questions with observational data is difficult because we observe the productivity only of workers who are hired. If referrals provide information about worker quality and firms (rationally) incorporate this information into their hiring decisions, hired referred workers may not perform better than hired nonreferred workers, even though the referral provides positive information about worker productivity.

Our experiments circumvent differential selection of referred and nonreferred workers into employment. By working in an online marketplace (oDesk), we were able to hire workers directly, allowing us to compare the performance of referred and nonreferred applicants, not just the workers a given firm chose to hire. The experiments took place between January and June 2013. We ran three experiments: the peer influence experiment, the team experiment, and the selection experiment. To recruit our samples for the peer influence and team experiments, we first hired experienced workers, asked them to complete a short task unrelated to the experimental tasks, and solicited referrals from those who complied. We then invited referred workers and a random sample of nonreferred workers to apply and hired all applicants who met our basic wage criteria. These two experiments were designed primarily to answer whether referred workers perform better because they are referred: because either (1) they work harder because they think their performance will affect their referrers' position at the firm or their relationship with their referrers (peer influence) or (2) they perform better when working directly with their referrers (team production).[1] Four months later, we conducted

periment. Sophie Wang provided excellent research assistance. Financial support from the Lab for Economic Applications and Policy at Harvard is gratefully acknowledged. Code is provided as supplementary material online.

[1] Peer influence leads referrals to work harder in Kugler's (2003) model because referrals face a psychic cost of exerting less effort than their referrers, while Dhillon, Iversen,

the selection experiment, designed to test whether referrals perform better than nonreferred workers even without on-the-job interactions with their referrers. We made job offers from a new firm to all referred and nonreferred workers (but not referrers) in the peer influence experiment.

We find that referrals do reveal positive information about worker quality independent of on-the-job interactions with referrers. In the selection experiment, referred workers exhibited substantially higher performance and lower turnover than did nonreferred workers even at a firm to which they had not been referred and at which their referrers did not work. Little to none of the information contained in the referral was otherwise observable to the employer through workers' resumes.

The peer influence experiment provides additional evidence that referrals contain information about worker quality. In this experiment, referred and nonreferred workers tested an airline flight website by answering questions about the site every other day over 12 days. Referrals in the peer influence experiment were randomized into two treatments. The nonmonitoring treatment was designed to minimize peer influence. Referrals in this treatment were told their referrers would never know their performance, and (after referring) referrers were told they would not be judged on the performance of their referrals. As in the selection experiment, referred workers in this treatment performed better and had less turnover than nonreferred workers, and these differences could not have been predicted from workers' observable characteristics. We also use data from this experiment to simulate a realistic hiring process and to show that we could have obtained misleading results if we had compared the performance only of applicants employers chose to hire.

The monitoring treatment of the peer influence experiment was designed to maximize peer influence. Each referrer in this treatment received an update on her referral's performance after each day of work. We implied to each referrer that her referral's performance and willingness to continue working for us would affect whether the referrer was promoted. Yet, we do not find that monitored referrals performed significantly better or had less turnover than nonmonitored referrals.

---

and Torsvik's (2012) and Heath's (2015) models suggest that referred workers work hard because if they perform poorly the firm will punish their referrers. This is similar to microfinance group lending wherein a worker's peers may pressure the worker to repay the loan (e.g., Bryan, Karlan, and Zinman 2015). While team production has not been emphasized as an explanation for hiring referrals in the economics literature, general research on team production implies that it may be an important benefit of referrals. For example, Bandiera, Barankay, and Rasul's (2013) model finds that when working in teams with their friends, workers are less likely to free-ride; Bandiera, Barankay, and Rasul (2005) find that workers are more able to cooperate with their teammates when their teammates are friends; and Costa and Kahn (2003) find that Civil War soldiers were less likely to desert when more of their unit were from their own birthplace.

The team experiment, however, does suggest that working directly with her referrer makes a referral more productive. Here, the task was to work with an assigned partner to create a single, shared slogan for a public service announcement (PSA). Each of the two partners was given a different information sheet containing a distinct criterion for the slogan (e.g., be exactly three words long). We asked the partners to use the chat box provided on the site to discuss the task and then each to submit the same slogan, which should have satisfied both criteria. Workers completed three such PSA tasks, each with a different partner. Importantly, each referral completed one task with her referrer and one task with another randomly chosen referrer. Referred workers performed substantially better when paired with their own referrers.

The tasks for all of our experiments were chosen to be similar to tasks that are common on oDesk, which has over 2.5 million workers (Horton 2017) and 35 million hours billed in 2012 (oDesk Corporation 2013). Nevertheless, an important caveat to our findings is that employer-employee relationships on oDesk are typically much shorter than those in offline labor markets. We discuss implications of these differences for the interpretation of our findings in Section V of the paper.

We see our results as reconciling the seemingly inconsistent findings from papers comparing the performance of referred and nonreferred workers. Among call center workers, Castilla (2005) finds that referred workers perform better than nonreferred workers, while among bank tellers, Blau (1990) finds that referred workers perform worse. Studying nine firms in three different industries, Burks et al. (2015) find that referred workers perform similarly to nonreferred workers on most metrics, though they have less turnover. We show that referrals contain information about worker quality, but that if employers utilize that information in the hiring process, hired referred workers could perform better than, worse than, or the same as nonreferred workers.[2]

Other papers directly test predictions of models in which referrals contain information about worker quality. Using firm data, Brown et al. (2012) find results consistent with these models: referred applicants are more likely to be hired and hired referrals have lower turnover and higher initial wages, though the wage advantage decreases over time. Dustmann, Glitz, and Schönberg (2011) find similar results using matched employer-employee data and ethnic minority groups to proxy for referrals. Inconsistent with these models, however, Pistaferri (1999) and Bentolila, Michelacci, and

[2] Burks et al. (2015) show that referred applicants are more likely to be offered jobs, all else equal. This is consistent with a model in which referrals contain information about worker quality, but also with other models like nepotism. The paper also finds that referred workers are more likely to accept job offers.

Suarez (2010) find that workers who find jobs through informal networks earn lower wages. Our paper adds to this literature by directly analyzing worker performance and by constructing a setting (the selection experiment) in which referrals' superior performance cannot result from on-the-job interactions with referrers. In an experiment, Beaman and Magruder (2012) find that, when told they will be paid on the basis of their referrals' performance, employees refer higher-performing workers. Our paper builds on this by showing that referred applicants perform better than nonreferred applicants. Finally, Heath (2015) finds that referrers' and referrals' wage changes are highly correlated, consistent with a peer influence mechanism.[3]

There is also closely related research that uses the oDesk platform. Stanton and Thomas (2014) carefully analyze oDesk agencies, formal groups of oDesk workers often formed through offline connections. Agency-affiliated workers pay a fraction of their earnings to their agency, and in return, their agency affiliation is listed on their resume. The paper finds that employers view agency affiliation as a signal that inexperienced workers are productive. Among inexperienced workers, employers are more likely to hire agency affiliates than unaffiliated workers, and they pay affiliates higher wages. Once workers have accumulated other signals of productivity (in particular, employer feedback scores), the importance of this signal declines. A related paper, Horton (2017), finds that oDesk employers value recommendations of whom to hire. Employers who randomly received recommendations about workers from oDesk itself were both more likely to hire these workers and more likely to hire anyone for their jobs. Yet, workers hired as a result of these recommendations were not more successful than other hired workers. There are a number of recent papers that use oDesk to learn about general features of labor markets (see, e.g., Ghani, Kerr, and Stanton 2014; Horton 2014, 2017; Pallais 2014; Stanton and Thomas 2014; Gilchrist, Luca, and Malhotra 2016; Lyons, forthcoming).

The remainder of the paper proceeds as follows. Section II describes the marketplace and our experimental designs. Section III analyzes whether referrals contain information about worker quality, Section IV examines whether referrals perform better because they were referred, and Section V discusses external validity. Section VI concludes the paper, discussing how these results could inform strategies to improve unconnected workers' labor market outcomes.

[3] A few papers suggest that firms prefer referrals for reasons other than improved productivity. Consistent with firms hiring workers' children as a favor to existing workers, Kramarz and Skans (2014) find that parents' wage growth drops dramatically exactly when one of their children is hired. Wang (2013) also finds evidence of nepotism in referrals. Holzer (1987) and Burks et al. (2015) find that hiring referred workers lowers recruiting costs.

II. Experimental Context and Recruitment Design

A. *Online Labor Market*

oDesk is an online labor market in which employers, mostly from the United States, hire independent contractors from all over the world for jobs that can be completed remotely. The jobs range from those that require significant skills such as computer programming or software development to less skill-intensive tasks such as data entry, internet research, or administrative support. Unlike Amazon's Mechanical Turk, another online marketplace commonly used in economics research, oDesk employers have complete discretion in whom they hire and they have real relationships with hired workers.

Employers post job listings describing their jobs and any required worker characteristics. They consider applicants' resumes when deciding whom to hire. (Figure 1 shows a sample oDesk resume from a worker not in the experiment.) These resumes contain information about workers' skills and qualifications as well as their past experience. The resumes list previous oDesk jobs, educational degrees, skills tests that workers have passed, and a 1–5 feedback score from previous employers. Employers can also choose to interview workers remotely before deciding whom to hire, though many employers do not.

Most jobs on oDesk, including all the jobs in this experiment, are hourly jobs (Pallais 2014). In these jobs, workers propose an hourly wage when they apply. Workers are then paid their set hourly wage for all hours worked, regardless of the output, though the employer can end the job and fire the worker at any time. Workers also post a desired hourly wage at the top of their resumes, which firms can observe.

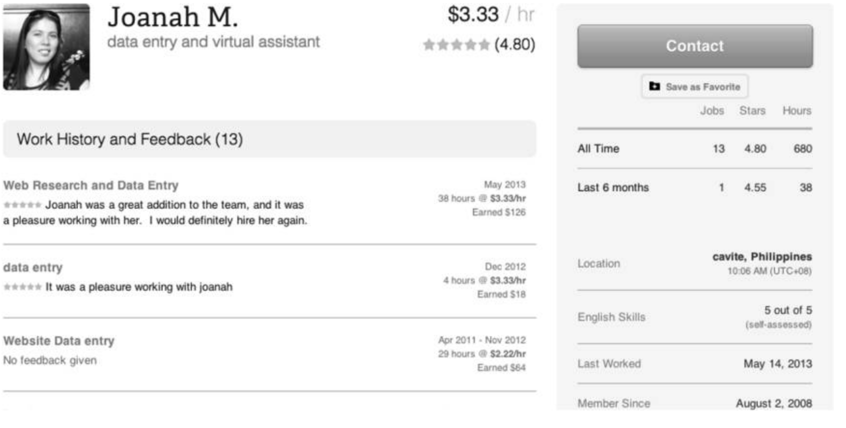

FIG. 1.—oDesk profile example

During the employment relationship, workers and employers communicate through the oDesk messaging system. They also use non-oDesk methods such as e-mail and Skype. oDesk allows employers to monitor workers' progress, similar to the monitoring that would be possible in an in-person environment. Workers log in to an oDesk application that shows employers when they are working. This application provides information about workers' keystroke volume and shows screen shots of the workers' computers, taken six times per hour.

Most workers state that they are available to work full-time (30+ hours per week), though others are available part-time or only a few hours per week.[4] In general, oDesk workers are relatively young and well educated and, among the lower-wage segment employed in these experiments, disproportionately likely to be female. Many workers have friends and relatives who also work on oDesk. Though there is at present no explicit referral mechanism on oDesk, employers can solicit referrals from their current workers and workers can recommend people they know to their employers. oDesk also has agencies, formal groups of oDesk workers often formed through offline connections (Stanton and Thomas 2014).

B. *Hiring Our Experimental Samples*

We hired workers for the peer influence and team experiments in the same way. (The sample for the selection experiment was a subset of the peer influence experiment sample.) We first invited a random sample of oDesk workers who (1) were from the Philippines, (2) listed an hourly wage of $5 or less on their resumes, (3) had earned $50 or more on oDesk, and (4) had an average job feedback score of 4 or higher to apply to our job. We eliminated workers with ratings below 4 because we wanted only referrals from workers we would actually hire; because most oDesk ratings are very positive, only 16 percent of workers who met our other criteria had ratings below 4. We included workers only from the Philippines because we wanted all workers in the team experiment to be able to communicate easily and be in the same time zone, and the Philippines is the most common country of residence for low-wage oDesk workers.[5] We told these workers very little about the task, only that we were hiring "for a variety of ongoing administrative support tasks of varying durations" and that we were looking for "diligent and highly-motivated individuals who are competent in the English language and interested in an ongoing relationship with our firm." We also told them that the position came with

[4] This statistic is from personal correspondence with John Horton and is based on calculations using oDesk administrative data.

[5] That the Philippines is the most common country of residence for low-wage oDesk workers comes from one of the authors' calculations using oDesk administrative data.

the possibility of promotion to managerial roles. We gave workers 48 hours to apply and then hired all workers who applied at an hourly wage of $3 or less.[6]

Original hires were asked to visit our website to initialize the job. The initialization step was intended to give workers some connection to our firm and to weed out the least responsive workers. (We fired the 5 percent of workers who did not initialize.) We then asked the workers who initialized to refer up to three other oDesk workers who were "highly qualified" and who they thought would "do a good job and be interested in an ongoing relationship with our firm." We did not provide workers with financial incentives for referring.[7] On each referral form we included questions about how well the referrer knew her referral, how often they interacted (remotely or in person), and how many people they knew in common. We also asked if they ever worked in the same room; since referrers might have more easily monitored or collaborated with referrals working in the same room, we eliminated from our sample any referral who ever worked in the same room as her referrer.

We invited to our job all referred workers who listed an hourly wage of $5 or less. (All workers who were referred were located in the Philippines.) We simultaneously invited to our job a random sample of oDesk workers from the Philippines with hourly wages of $5 or less.[8] We again gave workers 48 hours to apply. Referred workers were much more likely to apply to our job: 68 percent of referred workers applied versus only 6 percent of nonreferred workers. We then hired all referred and nonreferred workers who applied at an hourly wage of $3 or less.[9] We did

[6] We chose a $3 wage cutoff to minimize the cost of the experiment, while ensuring a sufficient sample size and a sample that was representative of the low-wage segment on oDesk. We initially contacted workers with wages of up to $5 as many workers are willing to work for wages below those listed on their resumes (Pallais 2014). For logistical reasons, we needed to hire workers at the same time. Because oDesk workers tend to remove their job applications if they do not hear back quickly, we gave workers 48 hours to apply. Prior experience suggested that 48 hours would maximize the size of the applicant pool.

[7] Online app. table 1 describes the characteristics of workers whom we asked to refer. It shows that workers who referred someone look somewhat more qualified than those who did not.

[8] We eliminated from the pool of both referred and nonreferred workers any workers who had already been invited as a potential referrer. We also eliminated from the team experiment anyone who had been invited in the peer influence experiment. As a result, referred and nonreferred workers in the team experiment look worse on observables than do referred and nonreferred workers in the peer influence experiment.

[9] We designed the recruitment process so that when referrers were submitting their referrals, they had no information about our actual tasks. The initialization step, e.g., was unrelated to the tasks themselves. From their own invitation to apply and from our request for referrals, referrers did know that we were hiring "for a variety of ongoing administrative support tasks of varying durations" and that we were looking for "diligent and highly-qualified individuals who are competent in the English language and interested in an ongoing relationship with our firm." However, all referred and nonreferred workers saw this same description on our job posting. Since referred workers had no private information

not tell original hires or their referrals anything about how they would be treated before the referral was made and the referred worker applied for the job. For example, original hires and referrals in both treatments of the peer influence experiment had the exact same information up until the time the referral was hired.

This recruiting process, used for both the peer influence and team experiments, produced an experimental sample with three types of workers: referred workers, nonreferred workers, and "referrers" (i.e., workers who made a successful referral). Figure 2 depicts this recruitment process. Workers who did not refer anyone or who referred a worker we did not hire performed a different, shorter task and are not included in any performance results. In the selection experiment, we made job offers to all referred and nonreferred workers from the peer influence experiment; no referrers were included. Figure 3 shows the recruitment of the selection experiment sample.

C. *Peer Influence Experiment Design*

The peer influence experiment was designed primarily to determine whether referrals work harder as a result of being referred because they think their performance and persistence will affect either their referrer's position at the firm or their relationship with their referrer. It also allows us to analyze whether referrals contain information about worker quality.

Panel A of online appendix table 2 describes the characteristics of the referred and nonreferred workers in the peer influence experiment. Referred workers, on average, had been on oDesk for about 18 months, and almost three-quarters had prior oDesk employment. Those who had been employed averaged over nine previous jobs and $1,382 in prior oDesk earnings. Nonreferred workers had been on oDesk slightly (insignificantly) longer but were less than half as likely to have previously been hired. Referred workers also had higher feedback scores from prior employers and were more likely to have passed oDesk tests. Despite being seemingly more experienced than nonreferred workers, referred workers posted wages on their resumes that were 15 percent lower than those posted by nonreferred workers, and they proposed significantly lower wages to our jobs. Recall that referred workers were also much more likely to apply to our job. This suggests that referrals may reduce recruiting costs by providing a way to identify workers with good resumes who are interested in the job.

We designed our task in this experiment to emphasize diligence because showing up to work and completing tasks in a timely manner are key determinants of success for low-skilled workers, both in more general labor

about the job before referring, in our context there is no scope for referrers to choose referrals with high worker-firm match quality.

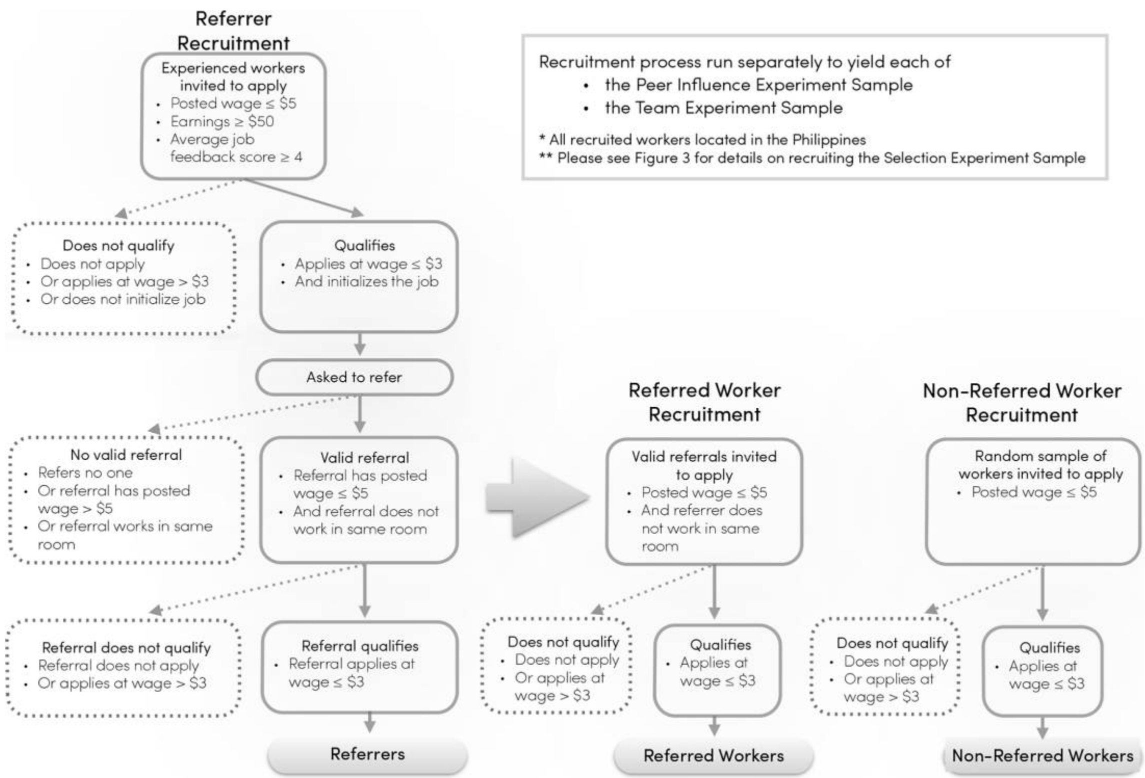

FIG. 2.—The recruitment process

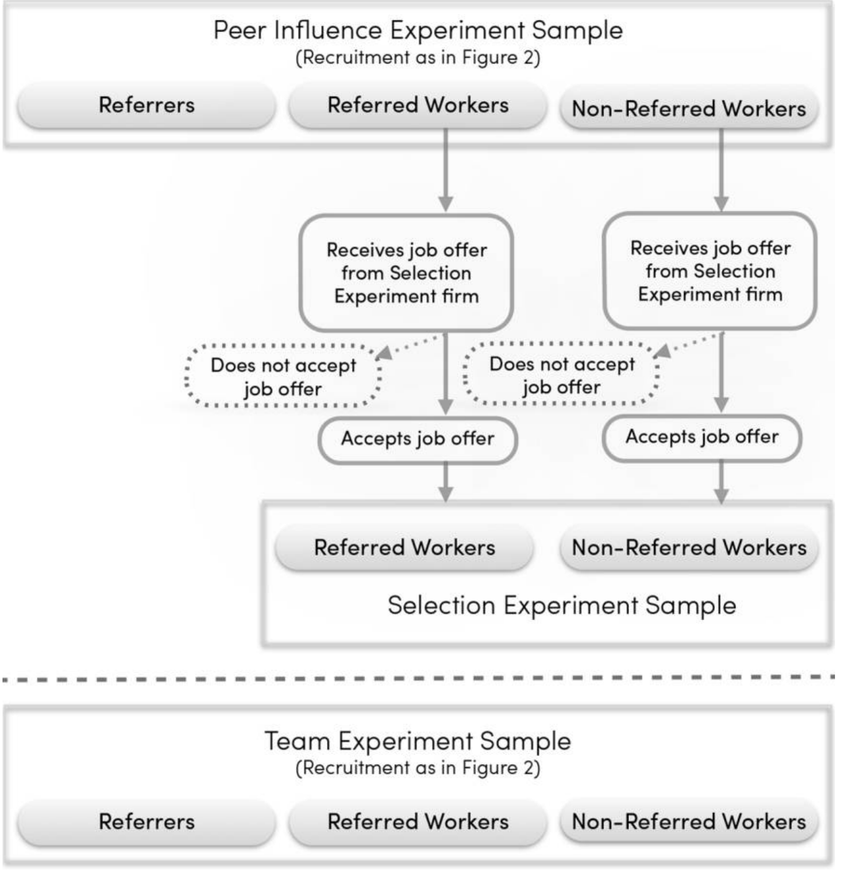

FIG. 3.—The experiment samples

markets and on oDesk (Holzer 1999; Regenstein, Meyer, and Hicks 1999; Pallais 2014). We also designed the task to measure worker turnover since decreased turnover is emphasized in the literature as a benefit of hiring referrals (e.g., Dustmann et al. 2011; Brown et al. 2012; Burks et al. 2015).

All referred and nonreferred workers in the experiment completed the same task. We told them they would be doing testing for an airline flight website and asked that they visit the site every other day for 12 days (six visits total), answering the questions on the site each day. For each worker on each day, the site displayed a table with a randomly generated set of 10 flights. Each flight was identified by a flight number and included a departure and arrival city, price, and number of available seats. Just below the flights table were six fill-in-the-blank questions (e.g., the flight number of the cheapest flight). The questions were the same each day, but the correct answers changed with the set of flights shown. Online appendix figure 1 displays a sample flights table followed by the questionnaire.

We told all referred and nonreferred workers to complete the task on the assigned day and asked, but did not require, that they complete each day's task by 11:00 a.m. Philippine time. We also informed all referred and nonreferred workers that we would send performance updates to a manager after each working day reporting (1) whether they submitted a response on the assigned day, (2) whether they submitted a response by 11:00 a.m. on that day, (3) whether they answered all the questions, and (4) the percentage of working days they had met each of these three performance criteria. Online appendix figure 2 shows an example performance report.

Referrers were randomized into the monitoring and nonmonitoring treatments. Each referred worker was assigned to the same treatment as her referrer. Online appendix table 3 shows that the randomization produced balanced samples between the treatment groups within both the referrer and referral samples. Out of 26 comparisons between the two treatment groups, only one difference is significant at the 10 percent level.[10]

The monitoring treatment was designed to facilitate monitoring of the referred worker by her referrer while the nonmonitoring treatment was designed to minimize peer influence. Referred workers in the monitoring treatment were told that their daily performance statistics would be sent to their referrer as well as to the manager. Referred workers in the nonmonitoring treatment meantime were explicitly told that their referrer would never see their performance statistics, only the manager would. The difference in performance and persistence between referred workers in these treatments is due to peer influence. The difference in performance between referred workers in the nonmonitoring treatment and nonreferred workers sheds light on whether referrals contain information about worker quality. However, even referred workers in the nonmonitoring treatment may have worked harder because they felt grateful for having been referred or faced informal pressures from their referrers.

Referrers worked on a different task. We wanted to employ them for the duration of their referrals' contracts, and we wanted them to understand the performance metrics we sent them about their referrals. Thus, we asked them to answer questions on a website every other day over the same 12-day period and we assigned them a soft deadline of 2:00 p.m. Philippine time for submitting. We did not, however, want the referrers to garner insights from their own task with which they could potentially help their referrals, so we had them work on a site that had a different

[10] While there are 28 comparisons in the table, by construction, there is no variation in prior experience or in having a feedback score among referrers.

log-in method, was focused on consumer products rather than flights, and asked a different set of questions.

To strengthen the treatment, we told all referrers before work began that they were being considered for a higher-paying management position. We implied to referrers in the monitoring treatment that whether they were promoted would depend on their referrals' performance.[11] Referrers in the nonmonitoring treatment were also informed of the management position but were assured that they would be "judged on their own merits" and that the performance of their referral would in no way influence the promotion decision. As promised, we sent the performance statistics of each referred worker in the monitoring treatment to her referrer. We also sent the referred and nonreferred workers' statistics to a manager we hired.

At the end of the task, we invited all referred and nonreferred workers to reapply to continue on the same project. We use this as an (inverse) measure of worker turnover. Each referred and nonreferred worker was told that the manager would receive an update on whether she accepted our offer to reapply. Referred workers in the monitoring treatment were told that this update would also go to their referrers while referred workers in the nonmonitoring treatment were explicitly told that their referrers would not see this information. To strengthen the treatment, when we invited referrers in the monitoring treatment to apply for the management position, we told them that we had just invited their referrals to continue on with their task and hoped their referrals would accept the invitation. We invited referrers in the nonmonitoring treatment to apply for the management position as well but made no mention at all of their referrals. This experimental design is summarized in panel a of figure 4.

D. Selection Experiment Design

The selection experiment was designed explicitly to determine whether referrals contain information about worker quality. Four months after the peer influence experiment, we measured the performance and persistence of referred and nonreferred workers in a job to which the referred workers had not been referred. We created a firm with a name, location, job posting, and writing style different from those in the peer influence experiment. We sought to hire the maximum possible number of referred

[11] All referrers were told that the management position would require being able to identify "high-ability workers interested in an ongoing relationship with our firm." When we told referrers in the monitoring treatment about the position, we also said that they would receive daily performance updates on their referrals "because we care about workers' performance." To make sure we were as truthful as possible, we hired some of these workers for management positions after the experiment.

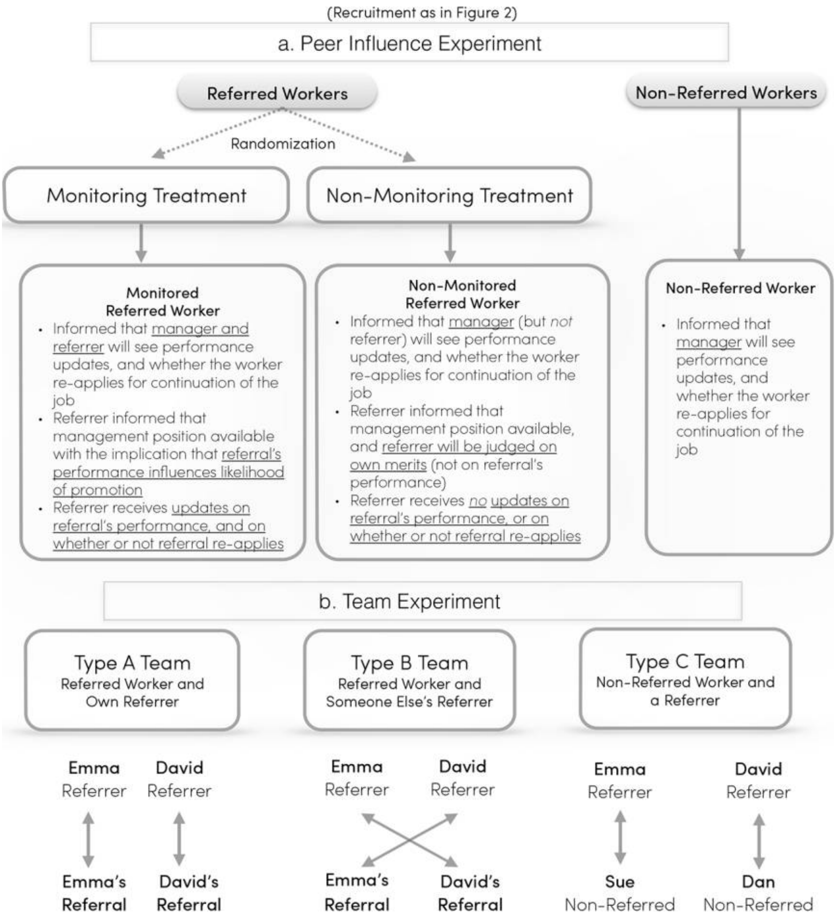

Fig. 4.—Treatments in the peer influence and team experiments

and nonreferred workers. We made direct job offers to all referred and nonreferred workers from the peer influence experiment and sent three reminders to accept to workers who had not yet responded. None of the referrers was contacted by this firm. Panel B of online appendix table 2 describes the characteristics of the referred and nonreferred workers who accepted our offer. (These characteristics were measured at the time we first contacted them for the peer influence experiment.)

Similarly to the peer influence experiment, workers who accepted the job offers were given a task that measured individual diligence over time. Workers were asked to visit the Twitter pages of three successful musicians and to answer a 10-question survey about those accounts every day for 5 consecutive days (Monday through Friday). We assured workers they needed no prior knowledge of Twitter and explained where to find the relevant information. Most of each day's task involved reporting on the Twitter activity of the artist from the day before. Although we asked workers to complete the task on the correct day, we also accepted retroactive submissions and automatically recorded the time of submissions. Online appendix figure 3 displays the questionnaire. After the last assigned day of work, we again invited workers to a continuation of the task and recorded whether they reapplied.

E. Team Experiment Design

The team experiment was designed to determine whether directly working with their referrers leads referrals to perform better (team production).[12] The task involved brainstorming, and we encouraged teamwork. Each worker was paired with three successive partners and asked to come up with a slogan for each of three different PSAs. We chose this task because there are many jobs on oDesk that ask low-skill workers to come up with advertisements, including jobs that specifically ask workers to create slogans. The first PSA was to encourage highway drivers to wear seat belts, the second was to encourage children to practice good dental hygiene, and the third was to encourage college students to get the flu vaccine. For each PSA, we asked the worker to use the chat box we provided on our site to communicate with her partner and to come up with a single slogan that both partners would submit through our online form. Online appendix figure 4 gives an example of what workers saw when they logged in to the team task site.

Though a worker could complete the task without her partner, the task was designed so that the best output necessitated teamwork. Each partner received a different sheet with information relevant to the PSA. For the first PSA, for example, one partner received information on seat belts' efficacy, while the other received information about highway drivers. The stated justification was that there was a lot of information to process and that by giving the partners different information, each partner would have to read only half as much. We told workers we wanted them to work with a partner to come up with their slogan because brainstorming is often more effective in teams.

Each information sheet contained a specific criterion we wanted the slogan to meet as well as a reason for that criterion. In the first round, for example, we told one partner that we wanted the slogan to be only

[12] Panel C of online app. table 2 shows the characteristics of referred and nonreferred workers in the team experiment. As in the peer influence and selection experiments, referred workers were more likely than nonreferred workers to have previously been hired and had higher feedback scores from prior employers, but they proposed significantly lower wages to our jobs.

three words long (so as not to distract drivers) and we told the other that we wanted the slogan to be in all capital letters (so drivers would be more responsive to it). In the second round, we told one partner to use an emoticon in the slogan (to make dental hygiene seem more upbeat) and the other to use the name of a real or fictitious person (since kids may respond to role models). In the third, we told each partner we wanted one of four specific words included in the PSA; one partner's word choices emphasized that getting the flu shot would be quick, and the other partner's word choices emphasized that flu shots are effective. When giving workers their information sheets, we told them only that the sheets would contain information, not that they would contain particular criteria for the slogans.

When workers submitted their slogans, we asked them also to answer a "team question": a multiple-choice question about the slogan. Each of the three PSA assignments had a different team question (what color sign the PSA should be printed on, what type of lettering the slogan should be written in, and where the PSA should be placed). This question had no correct answer, but partners were instructed to give the same answer.[13]

For comparison with the peer influence and selection experiments, we also collected measures of individual diligence. We monitored whether each worker logged in to the site and whether she submitted work. We also asked each worker an "individual question," the answer to which was in her own information sheet (e.g., the fraction of highway drivers who wear seat belts). Because workers were instructed that they should complete the task even if they could not make contact with their partner, workers should have logged in, submitted work, answered their individual question correctly, and used the criterion from their own information sheet in their slogan regardless of whom they were partnered with.

In the experiment, each referrer completed the three different PSA tasks as part of three different types of teams: (1) a type A team, in which she was paired with her own referral; (2) a type B team, in which she was paired with someone else's referral; and (3) a type C team, in which she was paired with a nonreferred worker. Panel b of figure 4 gives an example of these three team types. Each referred worker worked with her own referrer when her referrer was in a type A team and with someone else's referrer when her referrer was in a type B team. (When her referrer was in a type C team, she worked with another referred worker in the same position; results from this treatment are not presented.) Nonreferred workers worked with referrers for all three rounds; that is, they were always in type C teams.

[13] Because we wanted to measure how effectively workers worked with their partners, we strongly encouraged each worker to complete each PSA. In contrast to the peer influence experiment, in which we sent workers no reminders about the task, in the team experiment we sent two reminders about each PSA to each worker who had not already submitted work.

Comparing the performance of referred workers in type A and B teams provides the value of team production: how much better a referred worker performs when working with her own referrer than with someone else's referrer. Comparing the performance of workers in type B and C teams shows the difference between referred and nonreferred workers when both work with partners they do not know.

Because we thought worker performance might be correlated not just between partners but also among partners' partners, we placed workers into blocking groups. Each of the 47 blocking groups contained six referrers, their six referred workers, and two nonreferred workers. By definition, every worker in the blocking group partnered only with others in the same blocking group. In all analyses of the team experiment, we cluster standard errors by blocking group.[14] The placement into blocking groups was random, except that a referrer and her referral were always in the same group.[15] Within a blocking group, the ordering of the type of team workers participated in was random. And, within team type, when relevant, workers' assigned partners were also random.

In addition to measuring worker performance, we collected a proxy for worker enjoyment of the partnered task and willingness to continue working with each partner. After the worker submitted her last slogan, we asked, "In case we have more tasks like this in the future, which if any of the partners that you've worked with would you be interested in working with again?" Workers could select all, none, or a subset of their partners.

III. Referrals and Information about Worker Quality

We now examine whether referrals provide information about worker quality. First, we compare the performance and turnover of referred and nonreferred workers in the selection experiment. Then we compare nonmonitored referred workers and nonreferred workers in the peer influence experiment.

A. *Selection Experiment*

The selection experiment shows that referrals do contain information about worker quality: even working at a job for which they were not re-

[14] We do find evidence of learning from partners, supporting our decision to cluster by blocking group. We show in online app. table 4 that a team performed better when one of its members had previously been in a type A team, controlling for the current team type and the task number. Since the task order was random, this may suggest that when workers are in successful pairings, they learn how to do the task successfully and use that knowledge in subsequent tasks.

[15] As in the peer influence experiment, we hired all referred and nonreferred workers who met the selection criteria. However, only one randomly selected referral from each referrer and only 94 nonreferred workers were included in this experiment.

ferred at a firm with which their referrers were not affiliated, referred workers outperformed nonreferred workers and had less turnover.

Table 1 compares the outcomes of the referred and nonreferred workers in the selection experiment. First, we consider workers' likelihood of accepting a job. Panel A includes no controls. Consistent with the idea that hiring referred workers decreases recruiting costs, even among workers contacted for the selection experiment—who had previously participated in an experiment—referred workers were more likely to accept our job offer. While 51 percent of nonreferred workers accepted, 68 percent

TABLE 1
Performance and Persistence in the Selection Experiment: Base Group Is All Referred Workers

	Sample: All Referred and Nonreferred Workers	Sample: Referred and Nonreferred Workers Who Accepted Job Offer			
	Accepted Job Offer (1)	Submission (2)	On-Time Submission (3)	Accuracy (4)	Reapplication (5)
		A. No Controls			
Nonreferred	−.167***	−.106**	−.107**	−.035	−.195***
	(.047)	(.046)	(.048)	(.026)	(.059)
Observations	435	1,325	1,325	1,325	265
R^2	.029	.013	.012	.003	.046
		B. First-Order Controls			
Nonreferred	−.071	−.100*	−.098*	−.024	−.123*
	(.056)	(.057)	(.059)	(.033)	(.071)
Observations	435	1,325	1,325	1,325	265
R^2	.125	.079	.077	.048	.088
		C. First- and Second-Order Controls			
Nonreferred	−.046	−.114*	−.108	−.043	−.172**
	(.064)	(.064)	(.067)	(.036)	(.086)
Observations	435	1,325	1,325	1,325	265
R^2	.268	.236	.236	.186	.349
Base group mean	.678	.763	.735	.363	.815

Note.—Each column in each panel reports the results of a separate regression of the dependent variable (indicated by the column) on an indicator for being a nonreferred worker. Panel A includes no controls, panel B includes the first-order controls for worker characteristics listed in fn. 16, and panel C also includes second-order controls (the square of each nonbinary characteristic in fn. 16 and the interaction of each pair of characteristics in fn. 16). Observations in cols. 1 and 5 are workers, while observations in cols. 2–4 are worker-days. Regressions in col. 1 include all workers contacted for the selection experiment; regressions in the remaining columns include only workers who accepted the job offer. Standard errors are clustered at the worker level when observations are worker-days, and Huber-White standard errors are presented when observations are workers.

* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

of referred workers did. To determine how much of the information contained in the referral would have been observable to employers through workers' resumes, panels B and C of table 1 add control variables to the regressions in panel A. Panel B adds our main covariates: what we call first-order controls.[16] Panel C adds the squares of each of the (nonbinary) covariates and the interaction of each pair of covariates (our second-order controls) to the regressions. The table shows that the 17 percentage point difference in job acceptance is almost entirely explained by observable characteristics (in particular, prior oDesk experience and prior earnings in the marketplace), leaving only an (insignificant) 4.6 percentage point difference in acceptance rates once we add the first- and second-order controls.

Next, we consider the performance and persistence of workers who accepted the job offer. Measures of performance and persistence are regressed on a dummy for being a nonreferred worker (the base group is referred workers). We consider three measures of performance: (1) an indicator for submitting the day's work, (2) an indicator for submitting it on time, and (3) the fraction of questions answered correctly (accuracy). Unanswered questions are marked as incorrect. We also consider whether workers applied for a continuation of the task as a measure of persistence.

The table shows that referred workers submitted work on 76 percent of days, and the vast majority of these submissions were made on time. However, nonreferred workers were 11 percentage points less likely both to submit work and to submit the work on time. While 82 percent of referred workers reapplied for a continuation of the task, nonreferred workers were 20 percentage points less likely to do so. However, despite the fact that these coefficients are large and significant, the nonreferred dummy explains only a small share of variation in the outcome measures: just over 1 percent in the case of submission and on-time submission and approximately 5 percent in the case of persistence.

Panel a of figure 5 shows performance over the course of the experiment by worker type. Submission rates of referred workers were consistently higher than those of nonreferred workers. Both types of workers became less diligent over time, but diligence fell off much more for nonreferred workers. Thus, the performance gap between referred and nonreferred workers grew over the course of the job. Panel A of online appendix figure 5 shows that the other performance measures (on-time submission and accuracy) follow similar trends.

[16] These are an indicator for having any oDesk experience, total oDesk earnings, the number of previous oDesk assignments, oDesk feedback score, an indicator for not having a feedback score, the wage listed on the worker's resume, the number of days since joining oDesk, an indicator for having passed oDesk tests, an indicator for having a portfolio, the self-reported English skill level, an indicator for not reporting an English skill level, an indicator for being affiliated with an agency of oDesk workers, and the number of degrees listed on the resume.

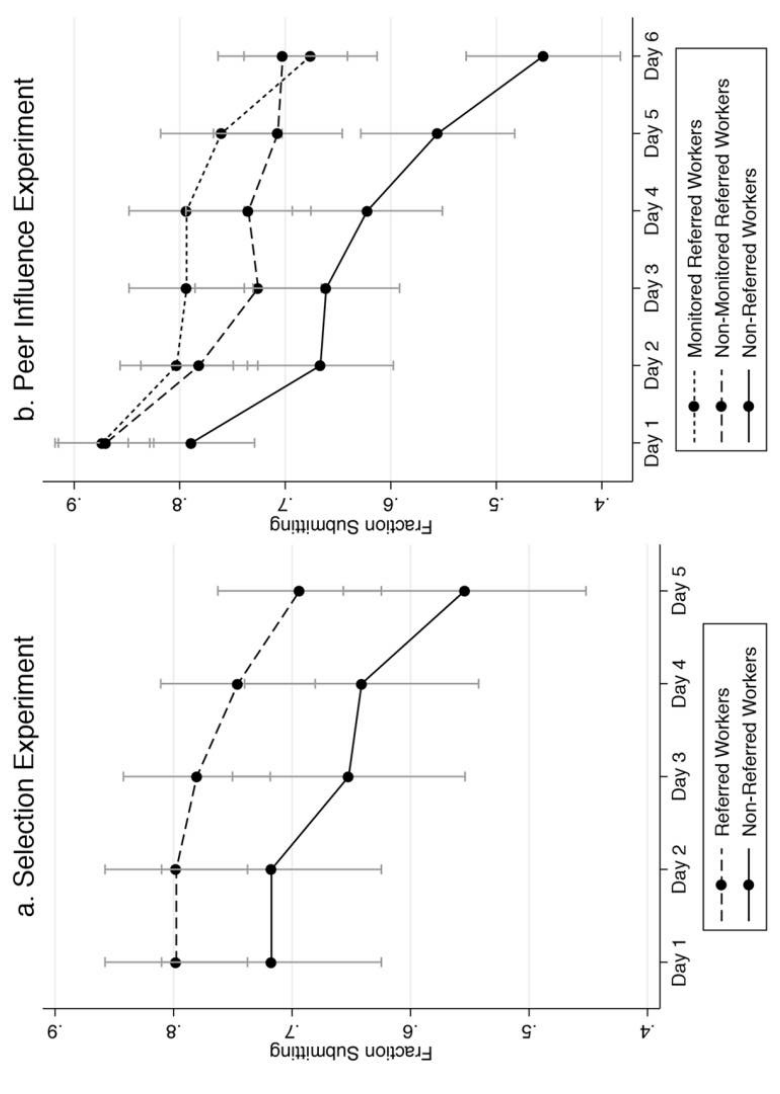

FIG. 5.—Submission rates by day. Error bars denote 95 percent confidence intervals.

Workers' resume characteristics are predictive of their performance and persistence: the proportion of variation explained increases to a quarter (for submission) and a third (for reapplication) when the first- and second-order controls are added. However, adding covariates does not change the coefficient on the referral dummy at all. This suggests that while the referral mostly contained observable information about workers' willingness to accept the job, most of the information contained in the referral about workers' performance and persistence was not otherwise observable through the workers' resumes. Panel A of online appendix table 5 displays the coefficients on the first-order controls from panel B of table 1. (Coefficients on the second-order controls are harder to interpret.) Unsurprisingly, the coefficients suggest that prior oDesk experience, more degrees, and passing oDesk tests—variables on which referred workers look better than nonreferred workers—are positively related to performance and persistence (though these coefficients are typically not significant). However, all else equal, the two characteristics that explain the most variation in performance are (1) having been on oDesk longer and (2) not being in an agency. Referred workers look worse on both these metrics.

B. *Peer Influence Experiment*

Next, we compare the performance and turnover of nonmonitored referred workers and nonreferred workers in the peer influence experiment. The results are very similar to those of the selection experiment. The main difference is that in the peer influence experiment, we also compare the performance of monitored and nonmonitored referred workers. We discuss this comparison in Section IV.

Each column of table 2 presents the results of regressing an outcome on an indicator for being a monitored referred worker and an indicator for being a nonreferred worker. (The omitted group is nonmonitored referred workers.) We use the same performance and persistence metrics as in the selection experiment: submission, on-time submission, accuracy, and reapplication.[17] Panel A includes no controls, panel B includes first-order controls, and panel C includes first- and second-order controls.

Referred workers performed better than nonreferred workers. Nonmonitored referred workers were 13 percentage points more likely to submit, 8 percentage points more likely to submit on time, and 23 percentage points more likely to reapply for the job than were nonreferred workers. Panel b of figure 5 shows that, as in the selection experiment,

[17] Two of the three performance metrics we consider are metrics the workers were told the manager would see daily: an indicator for submitting any response on a given day and an indicator for submitting the response by 11:00 a.m. Workers were also told that the manager would see whether they answered all questions, but we exclude this metric from our analysis since 99.8 percent of submissions were complete.

TABLE 2
PERFORMANCE AND PERSISTENCE IN THE PEER INFLUENCE EXPERIMENT:
BASE GROUP IS NONMONITORED REFERRED WORKERS

	Submission (1)	On-Time Submission (2)	Accuracy (3)	Reapplication (4)
	A. All Days, No Controls			
Monitored referred	.036	.053	.034	−.032
	(.042)	(.047)	(.039)	(.030)
Nonreferred	−.132***	−.079*	−.101**	−.225***
	(.042)	(.045)	(.039)	(.038)
Base group mean	.757	.563	.640	.953
Observations	2,610	2,610	2,610	435
R^2	.027	.013	.020	.085
	B. All Days, First-Order Controls			
Monitored referred	.020	.038	.014	−.035
	(.042)	(.046)	(.040)	(.034)
Nonreferred	−.115**	−.080*	−.095**	−.193***
	(.047)	(.048)	(.043)	(.044)
Base group mean	.757	.563	.640	.953
Observations	2,610	2,610	2,610	435
R^2	.075	.061	.063	.156
	C. All Days, First- and Second-Order Controls			
Monitored referred	.004	.045	.002	−.044
	(.041)	(.047)	(.039)	(.037)
Nonreferred	−.135***	−.067	−.100**	−.196***
	(.049)	(.052)	(.046)	(.052)
Base group mean	.757	.563	.640	.953
Observations	2,610	2,610	2,610	435
R^2	.181	.139	.163	.264
	D. Last Day Only, First- and Second-Order Controls and Daily Performance Controls			
Monitored referred	−.058	.018	−.046	−.053
	(.048)	(.057)	(.042)	(.038)
Nonreferred	−.172***	−.102*	−.103**	−.149***
	(.053)	(.056)	(.046)	(.050)
Base group mean	.703	.500	.600	.953
Observations	435	435	435	435
R^2	.614	.506	.622	.434

NOTE.—Each column in each panel reports the results of a separate regression of the dependent variable (indicated by the column) on an indicator for being a referred worker in the monitoring treatment and an indicator for being a nonreferred worker. As in table 1, panel A includes no controls, panel B includes the first-order controls for worker characteristics listed in fn. 16, and panel C includes first- and second-order controls. Regressions in panels A, B, and C include observations on all 6 days of work. Regressions in panel D are limited to observations on workers' last day of work and include first- and second-order controls as well as daily performance controls; each of cols. 1–3 includes controls for the worker's performance as measured by the dependent variable on each of the first 5 days of work. Column 4 includes controls for each of the three performance measures on each of the 6 days. Observations in cols. 1–3 are worker-days; observations in col. 4 are workers. Standard errors are clustered at the worker level when observations are worker-days, and Huber-White standard errors are presented when observations are workers.
* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

the submission gap between referred and nonreferred workers grew with time.[18]

Observable characteristics from workers' resumes explain a lot of the variation in outcomes, but they do not diminish the predictive power of the referral. The proportion of variance in performance explained increases from approximately 2 percent to 15–20 percent when all the controls are added, but the coefficient on the nonreferred dummy remains constant. Panel B of online appendix table 3 shows that the coefficients on the first-order controls are similar to the coefficients on these controls in the selection experiment regressions.

These results suggest that referrals contain information about worker performance that is not present in workers' resumes. In addition to using workers' resumes, firms could gain information about worker quality through interviews or a job test, both of which are costly.[19] While we do not know what information firms would gain through interviews, we can approximate the information that might be gained from a job test using workers' initial job performance. Panel D of table 2 shows that the referral still has predictive power for worker performance on the last day of the contract, conditional on worker performance on all prior days. Panel D replicates panel C, limiting the observations to the last day of the contract. Regressions in columns 1–3 now additionally control for the worker's performance (on the same metric as measured by the dependent variable) on each of the first 5 days. All differences in performance between referred and nonreferred workers remain large and significant.

The referral also provides information about worker persistence at the firm above and beyond the information provided by the worker's performance throughout the full contract. Column 4 of panel D adds controls for each of our performance measures (submission, on-time submission, and accuracy) on each of the 6 days. Even controlling for all our performance measures on all days, referred workers were 15 percentage points more likely than nonreferred workers to want to continue with the firm.[20]

The results suggest that referrals provide important information about worker quality. Even when referred workers were not monitored by their referrers, they performed much better than nonreferred workers and were more eager to continue with the firm. This information was not present on workers' resumes or in their performance on the majority of the contract.

[18] Panel B of online app. fig. 5 shows that the other performance measures (on-time submission and accuracy) follow similar trends.

[19] Even when firms undertake interviews, firms have considerable uncertainty about worker productivity when hiring (e.g., Autor and Scarborough 2008).

[20] Unreported coefficients show that workers who performed better were more likely to want to continue with the firm.

C. *Heterogeneity by Referral Type*

The above analysis suggests that referrals contain information. Here, we find that some referrals contain more information than others. In particular, referrals made by high-performing referrers and referrals of workers with strong ties to their referrers are particularly informative.

Using data from the peer influence experiment, column 1 of table 3 shows that a referrer's performance is a strong predictor of her referral's performance. This is not just a result of the referrer and her referral facing common shocks. The referrer's performance in the peer influence experiment is a strong predictor of her referral's performance in the selection experiment 4 months later (col. 2, table 3).

Some of this can be accounted for by observable characteristics. Online appendix table 6 shows that workers with better observable characteristics refer workers who also have better observables. Controlling for the referred worker's observable characteristics in the regression of referral performance on referrer performance reduces the point estimate on referrer performance. Nonetheless, the referrer's performance remains an important predictor of her referral's performance. This suggests that higher performers refer workers who perform better than would even be expected on the basis of their observable characteristics. It also suggests that not all referred workers are predicted to outperform nonreferred workers. In both the selection and peer influence experiments, these results suggest that referrals from the worst-performing 20 percent of referrers are predicted to underperform nonreferred workers, and in fact, they do.

We turn now to the relationship between referrers and their referrals. Online appendix table 7 shows the distributions of the three relationship variables we have from referrers' reports at the time of the referral. Referrers tended to refer workers they were close to. Most reported knowing their referrals "extremely well" (6 on a scale of 1–6), while only 1 percent said they knew their referral "hardly at all" (1 on the same scale). According to referrers, 32 percent of referrals interacted with their referrers more than once a day (in person or remotely) and another 19 percent interacted about once a day; meanwhile, only 7 percent interacted once a month or less. Just under half of referred workers knew 20 or more people in common with their referrers.

Because the relationship variables are positively correlated and predict performance in the same way, we build an index of relationship strength and for parsimony focus here on the resulting estimates.[21] In the final col-

[21] In building the index, we first create dummy variables for reportedly knowing the referred worker well (responding more than 3 on a scale of 1–6 when asked how well she knew the referred worker), interacting with the referral at least once a day, and knowing at least 30 people in common. Our relationship index is defined as the standardized sum of these three binary variables. We exclude the five referred workers whose referrers did not answer all the relationship questions at the time of the referral.

TABLE 3
HETEROGENEITY IN REFERRAL PERFORMANCE, ALL EXPERIMENTS;
DEPENDENT VARIABLES INDICATE REFERRED WORKERS' PERFORMANCE

	SUBMISSION RATE		SUBMISSION		SAME SLOGAN
	Peer Influence Experiment (1)	Selection Experiment (2)	Peer Influence Experiment (3)	Selection Experiment (4)	Team Experiment (5)
	A. No Controls				
Referrer's submission rate, peer influence experiment	.421***	.342***			
	(.066)	(.084)			
Relationship strength index			.030	.015	.044**
			(.022)	(.026)	(.021)
Dependent variable mean	.775	.763	.775	.760	.520
Observations	255	173	1,512	855	560
R^2	.192	.115	.006	.001	.007
	B. First-Order Controls				
Referrer's submission rate, peer influence experiment	.392***	.194**			
	(.068)	(.084)			
Relationship strength index			.036*	.027	.046**
			(.021)	(.023)	(.020)
Dependent variable mean	.775	.763	.775	.760	.520
Observations	255	173	1,512	855	560
R^2	.266	.272	.098	.186	.051

NOTE.—Each column in each panel reports the results of a separate regression in which the dependent variable is indicated by the column. All dependent variables indicate referral performance. In cols. 1 and 2, observations are referred workers. In these columns, referred workers' average performance over the course of the indicated experiment is regressed on their referrer's average submission rate in the peer influence experiment. Huber-White standard errors are in parentheses. In cols. 3 and 4, observations are worker-days and standard errors are clustered by worker. In col. 5, observations are at the worker-PSA level and standard errors are clustered by blocking group. In each of cols. 3–5, the dependent variable is regressed on an index for the strength of the referrer-referral relationship. This index is defined in Sec. III.C of the text and has mean zero and standard deviation one. Regressions in panel A include no controls, while regressions in panel B include the first-order controls for worker characteristics listed in fn. 16.
* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

umns of table 3, we regress referral performance in the different experiments on this index. In each experiment, the coefficients suggest that referrals with stronger ties to their referrers performed better. These coefficients actually increase slightly when we add controls for worker characteristics (panel B). The reason is that referrals with stronger ties to their referrers look worse on paper: they have lower earnings, have been on oDesk for less time, and have fewer educational degrees. Conditional on observable characteristics, a referred worker with a one standard deviation stronger relationship with her referrer was approximately 4 percentage points more likely to submit work in the peer influence experiment, 5 percentage points more likely to submit the same slogan as her partner in the team experiment, and (an insignificant) 3 percentage points more likely to submit work in the selection experiment.

These results are consistent with the idea that when workers refer people they know well, they choose workers who do not look as good on paper but who perform well in ways that would not be predicted by their observable characteristics.

D. *Potential Bias from Employers' Hiring Decisions*

To test whether referrals provide information about the expected performance of job applicants, we hired all applicants who met our basic hiring criteria. Here, we use our experimental data to simulate how our comparisons between referred and nonreferred workers might have been biased had we observed the performance only of workers an employer chose to hire.

Using data from the peer influence experiment, we first simulate which workers employers would hire if they observed only the characteristics on workers' resumes; we then simulate whom employers would hire if they additionally observed which workers had been referred. For comparison, we also show the characteristics of workers hired if employers observed only workers' referral status and no other characteristics. We assume that employers want to maximize the fraction of workers who submit a response on a given day and that they know the relationship between demographics and referral status and performance.[22] Employers predict each applicant's performance using the information they observe and then hire the half of the applicant pool with the best predicted performance.

Table 4 shows the results of the simulations. Results in the first row simulate hiring under the assumption that employers see only workers' resumes, not who was referred. The second row simulates hiring under the assumption that employers see only workers' referral status, so they hire a random sample of referred workers. Finally, the third row simulates

[22] In practice, an employer may prefer to hire a referred worker over a nonreferred worker who is predicted to perform slightly better either as a source of compensation to an existing employee or because the referred worker is predicted to persist longer at the firm. For simplicity and clarity, we abstract away from any such considerations here.

TABLE 4

Simulated Hiring Peer Influence Experiment: Assuming Top 50 Percent of Applicants Hired

	A. Fraction Hired (%)		B. Measure of Observables (%)		C. Actual Submission Rate (%)			
	Referred Applicants (1)	Nonreferred Applicants (2)	Hired Referred Workers (3)	Hired Nonreferred Workers (4)	Hired Referred Workers (5)	Hired Nonreferred Workers (6)	All Hired Workers (7)	Difference (8)
Observe characteristics only	58	39	80	78	83	70	78.5	13**
Observe referral status only	85	0	74	NA	77	NA	77.5	NA
Observe characteristics and referral status	79	9	77	83	79	76	79.1	3
Applicant pool average			74	68	77	63	71	15***

Note.—Each row presents the results of a separate hiring scenario. In each scenario, employers use the available characteristics to predict workers' performance (likelihood of submitting work) and hire the 50 percent of workers with the highest predicted performance. The first row simulates hiring under the assumption that employers observe only workers' resume characteristics, but not their referral status. To calculate a given worker's predicted performance in this scenario, the performance of all other workers (excluding herself) are regressed on their resume characteristics listed in fn. 16. The estimated coefficients are then used to predict the excluded workers' own performance. Throughout the table, this prediction of performance—based on a worker's observable characteristics alone—is used as the measure of observable characteristics in panel B. The second row assumes that employers observe only referral status, so they hire a random sample of referred workers such that 50 percent of the workforce is hired. The third row simulates hiring assuming that employers observe workers' resume characteristics and referral status. To calculate a given worker's predicted performance here, the performance of all other workers (excluding herself) are regressed on their resume characteristics and referral status and the resulting coefficients are used to predict the worker's performance. For each scenario, panel A presents the fraction of referred and nonreferred workers hired. Panel B presents the estimated probability, based on their observable characteristics alone, that hired referred and nonreferred workers submit work. Panel C presents the actual submission rate of the hired workers. Column 8 provides the difference in average submission rates of the referred and nonreferred workers hired under each scenario.

** Significant at the 5 percent level.

*** Significant at the 1 percent level.

hiring under the assumption that employers observe workers' resume characteristics and who was referred.

If employers observed only workers' resume characteristics, a higher fraction of referred (58 percent) than nonreferred (39 percent) workers would be hired (panel A). However, if employers also observed who was referred, the fraction of referred applicants who would be hired jumps to 79 percent; in the meantime, only 9 percent of nonreferred applicants would be hired. If employers observed only referral status, they would hire 85 percent of referred workers and no nonreferred workers.

Panel B displays the summary measure of the hired workers' observable characteristics. It shows that when employers observe workers' resumes as well as who was referred, hired nonreferred workers are positively selected on observables relative to hired referred workers.

Panel C shows the actual submission rates of the workers hired in each scenario. Compared to hiring at random, both (1) hiring using only observable characteristics and (2) hiring using only referral status substantially improve the performance of hired workers. (Hiring using these strategies relative to hiring at random improves the performance of hired workers by 7 and 6 percentage points, respectively.) Observing both referral status and observable characteristics brings slightly larger gains in performance than using either in isolation. These results suggest that referrals might provide a way for firms to reduce recruiting costs. Given that much of the gain from using workers' characteristics in hiring could be obtained from using referral status alone, if collecting information on workers' characteristics is costly, employers might choose to forgo collecting these characteristics in favor of using referrals.

The table also shows that if employers did not observe who was referred, hired referred workers would be substantially (13 percentage points) more likely to actually submit work than nonreferred workers. However, this difference would be only 3 percentage points (and statistically indistinguishable from zero) if employers also observed who was referred. This suggests that if we had observed only the performance of hired workers and did not observe all the characteristics employers used in their hiring decisions, we might have mistakenly concluded that referrals contained little to no information about worker performance.

IV. Effect of On-the-Job Interactions with Referrers

We now consider whether being referred actually makes referred workers more productive. First, we consider whether referrals work harder because they believe their performance will affect their relationship with their referrer or their referrer's position at the firm (peer influence). Second, we consider whether referrals perform better when working directly with their referrer (team production).

A. *Peer Influence*

The peer influence experiment shows that peer monitoring does not have a detectable effect on performance.

Anecdotal evidence suggests that referred workers in the monitoring treatment were, in fact, monitored by their referrers. Many referrers in this treatment replied to our daily performance reports and indicated a strong interest in their referrals' performance. They often apologized when their referrals had not completed the task on the preceding day or had not completed it by the soft deadline, and they assured us they would encourage their referrals to do better on subsequent days. Yet table 2 shows that while the coefficients indicate that monitored referred workers performed better than nonmonitored referred workers, these differences are much smaller than the differences between nonmonitored referred workers and nonreferred workers and are never statistically significant.[23] The negative (though again insignificant) coefficient on the monitored referred worker dummy in column 4 suggests that monitored referred workers were, if anything, slightly less likely to be interested in continuing with the firm, perhaps because they disliked being monitored.

Panel b of figure 5 sheds some light on how the performance of monitored and nonmonitored referred workers evolved over time. On the first day of work, before any performance reports had been sent, monitored and nonmonitored referred workers performed equivalently. The graph suggests that peer influence may have stemmed the drop-off in performance in days 2, 3, and 4 among monitored referred workers, though the differences between monitored and nonmonitored referred workers on those days is not significant. By day 6, however, monitored referred workers were no more likely than their nonmonitored counterparts to submit work.

Overall, we do not find robust evidence in favor of peer influence, though we cannot rule out the presence of peer influence, particularly at the beginning of the contract.

B. *Team Production*

The team experiment shows that referred workers perform better when working directly with their referrers. In particular, referred workers performed much better when working with their own referrer than with a randomly selected referrer they did not know.

[23] Using seemingly unrelated regression, we calculate the variance-covariance matrix between the coefficients in these three performance regressions and test the hypothesis that all three monitored referred coefficients are equal to zero. We are unable to reject this hypothesis.

We first consider the effect of team type on measures that do not rely on teamwork but may be indicative of individual diligence. These are indicators for logging in to our site to see the given PSA task, submitting work, correctly answering the question about their own individual reading, and including the criteria from their own information sheets in their slogans.[24]

In panel A of table 5, each measure of individual diligence is regressed on an indicator for being in a type A team (a referred worker paired with her own referrer) and an indicator for being in a type C team (a nonreferred worker paired with a referrer). The omitted group contains workers in type B teams (referred workers paired with someone else's referrer). Thus, the coefficients on the type A dummy indicate how much better referred workers perform when paired with their own referrer than when paired with someone else's referrer; the coefficients on the type C dummy indicate how much worse nonreferred workers perform than referred workers when both are paired with someone else's referrer. Each observation is a partner pair, but in these diligence measures, we consider only referred and nonreferred workers. Referrers' performance does not vary significantly across team types. First- and second-order controls for both partners' observable characteristics are included throughout.

On average, referred workers performed well on these diligence measures. Similarly to our previous results, nonreferred workers were less diligent than referred workers, even when neither group was working with a partner they previously knew.

Referred workers were 5 percentage points more likely to submit work and to correctly answer the question about their own reading when they were paired with their own referrer than when paired with someone else's referrer. Given that these are measures of diligence more than teamwork, this could suggest that referred workers exerted more effort when working with their referrer. The reason may be that, in this case, their performance affected their referrers' output. Alternatively, it could result from peer influence if working together made it easier for referrers to monitor their referrals.

Panel B compares team performance by team type. Observations are again at the partner-pair level. Referred workers did particularly well when working with their referrers. For example, referred workers were substantially (29 percentage points) more likely to answer the team question the same way when working with their own referrers than when paired with referrers they did not know; of the type A teams that both submitted responses, only 6 percent failed to submit the same response to the team

[24] If a worker did not answer the question about her reading, she is marked as not answering it correctly. Similarly, if she did not submit a slogan, she is marked as not including her own criterion in the slogan.

TABLE 5
INDIVIDUAL DILIGENCE AND TEAM PERFORMANCE TEAM EXPERIMENT: BASE GROUP IS REFERRED WORKERS PAIRED WITH SOMEONE ELSE'S REFERRER (Type B Teams)

	A. Individual Diligence			
	Logged In (1)	Submitted (2)	Individual Question Correct (3)	Own Criterion in Slogan (4)
Referred worker when working with own referrer (type A)	.018	.046**	.053*	.004
	(.018)	(.018)	(.030)	(.035)
Nonreferred worker when working with referrer (type C)	−.082	−.129**	−.159***	−.039
	(.053)	(.055)	(.054)	(.057)
Base group mean (type B)	.883	.837	.755	.440
Observations	846	846	846	846
R^2	.419	.381	.294	.180
	B. Team Performance			
	Both Submitted (1)	Team Question Matches (2)	Same Slogan (3)	Same Slogan and Both Criteria (4)
Referred worker and own referrer team (type A)	.099***	.287***	.372***	.103***
	(.024)	(.030)	(.034)	(.025)
Nonreferred worker and referrer team (type C)	−.122**	−.062	−.023	.004
	(.058)	(.054)	(.055)	(.036)
Base group mean (type B)	.730	.496	.337	.142
Observations	846	846	846	846
R^2	.312	.317	.313	.157

NOTE.—Each column in each panel reports the results of a separate regression of the dependent variable indicated by the column on indicators for being in a type A team and for being in a type C team. Observations in panel A are at the worker-PSA level; only referred and nonreferred workers (not referrers) are included. Observations in panel B are at the team-PSA level. All regressions include the first- and second-order controls for worker characteristics listed in fn. 16. Standard errors are clustered at the blocking group level.
* Significant at the 10 percent level.
** Significant at the 5 percent level.
*** Significant at the 1 percent level.

question. The results are consistent across team performance metrics. Column 3 shows similar results for submitting the same slogan. Only about one-third of type B teams submitted the same slogan, while type A teams were more than twice as likely to do so.[25] Online appendix A shows that

[25] One hypothesis is that firms could replicate the benefit of team production that comes from referrals by creating teams of workers with similar observable characteristics. However,

in addition to performing better, type A teams enjoyed their task more, spent more time on the task, and communicated more. They performed better even conditional on time spent and communication.

V. External Validity

Completing these experiments in an online labor market provides two major benefits. First, it allows us to observe the performance and persistence of workers without the filter of firms' hiring decisions. Second, it allows us to vary parameters of the jobs to cleanly identify why referred workers perform better and have less turnover than nonreferred workers. As with any field experiment we might run, however, the results of this experiment come from one particular labor market, in this case an online labor market.

The types of tasks in our experiments are not uncommon in offline labor markets. Autor, Levy, and Murnane (2003) classify tasks into five categories, now prevalent in the skills literature: expert thinking, complex communication, routine cognitive tasks, routine manual tasks, and nonroutine manual tasks. Our selection and peer influence experiments center on routine cognitive tasks such as basic computations and data entry.

Routine cognitive tasks are prevalent in offline labor markets, especially among workers with a high school diploma or some college. Autor et al. (2003) define a composite measure of routine cognitive tasks, which they map to census occupations using O*Net data. They find that occupations in office and administrative support are particularly heavy in routine cognitive tasks; examples include cashiers, customer service representatives, and tellers.

We think that the principal difference between oDesk and offline labor markets is the incentives workers face. Because oDesk jobs are typically shorter than offline jobs, oDesk workers are often less tied to any particular employer than are workers in other labor markets. Prior to our experiment, the average job taken by the referrers in our sample paid $237 and lasted 81 working hours. If oDesk workers are less concerned about their reputations with their employers than are most workers in offline labor markets, this could lead referrals to contain less information about worker

we do not find evidence that teams in which partners had similar characteristics perform better. We create indicators for whether both partners were of the same gender (using workers' names and honorifics), whether they lived in the same city, and whether they had previously worked at the same oDesk firm; we also measure the difference between the partners' wages. Partners in type A teams look more similar on each of these dimensions than do partners in type B teams. None of these similarities positively predicts performance; nor does including measures of them in the regressions affect the estimated effect of working with one's own referrer.

quality and on-the-job interactions with referrers to be less effective in improving worker performance.

Referrers in our experiments were not provided compensation to provide referrals or to provide high-quality referrals. They may have received a social benefit or felt a warm glow from helping a friend find employment (e.g., Beaman, Keleher, and Magruder 2013). But their incentive to make high-quality referrals was implicit: by making high-quality referrals they could improve their relationship with our firm. We did try to provide some incentives for workers to care about their relationship with our firm by implying that if they performed well, they could have a long-term relationship with us. Nonetheless, this working relationship was still far less long-sighted than working relationships in most labor markets. The fact that referrals still contained positive information about worker quality despite referrers' relatively weak incentives to refer high-quality workers suggests that referrals are likely to contain positive (and perhaps even more positive) information about worker quality in other labor markets.

Relatedly, if referrers are less concerned about their reputations with their employers on oDesk, they may exert less pressure on their referrals to perform well, weakening the effect of peer influence. Since we were aware that peer influence might not be as strong a motivator on oDesk as in other labor markets, we aimed explicitly to maximize the effect of peer influence in the monitoring treatment of the peer influence experiment. That we find very limited effects of peer influence, then, suggests that peer influence is likely not an important mechanism in this context. Nonetheless, peer influence may still be important in other labor markets, especially in labor markets in which referrers care more about ongoing relationships with firms.

VI. Conclusion

The use of social connections is ubiquitous in the labor market. More than half of jobs are found through informal connections, and firms are more likely to hire referred than nonreferred applicants, all else equal. This suggests that workers without social connections may be disadvantaged in the labor market (e.g., Montgomery 1991; Calvo-Armengol and Jackson 2004). This paper examines why firms prefer to hire referred workers: do referrals allow firms to hire more productive workers because they signal worker quality or because being referred actually makes workers more productive?

Understanding why firms prefer to hire referred workers can inform potential policy responses that may help unconnected workers. For example, if referrals provide information about worker quality, then providing unconnected workers with other ways to signal their abilities may improve their labor market outcomes (as in Pallais 2014). On the other hand, if team production actually causes referrals to be more productive, information approaches may not help unconnected workers. Nepotism may also be harder to eliminate.

We find strong evidence that on-the-job interactions between referred workers and their referrers lead referrals to perform better. While we do not find evidence of peer influence, our results suggest that team production is an important benefit of referrals. However, we also find strong evidence that referrals contain information about worker performance and turnover. In our context, referrals contain information about general productivity. In other contexts, referrals might also signal that a worker is a particularly good match for a given firm or job. While this explanation is precluded in our experiments because referrers did not have information about the job they were referring for, it could be important in other settings.

From our experiments, we learn that referrals made by high performers and referrals of workers with strong ties to their referrers were particularly informative. Yet we do not know why, that is, whether referrers actively choose referrals they know will perform well (as in Beaman and Magruder 2012) or whether these results obtain simply because productive workers have productive friends (as in Montgomery 1991). Understanding why referrals contain so much information about worker quality is an important question for future research.

References

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Q.J.E.* 118 (4): 1279–1333.

Autor, David H., and David Scarborough. 2008. "Does Job Testing Harm Minority Workers? Evidence from Retail Establishments." *Q.J.E.* 123 (1): 219–77.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2005. "Social Preferences and the Response to Incentives: Evidence from Personnel Data." *Q.J.E.* 120 (3): 917–62.

———. 2013. "Team Incentives: Evidence from a Firm-Level Experiment." *J. European Econ. Assoc.* 11 (5): 1079–1114.

Beaman, Lori, Niall Keleher, and Jeremy Magruder. 2013. "Do Job Networks Disadvantage Women? Evidence from a Recruitment Experiment in Malawi." Working paper, Univ. California, Berkeley.

Beaman, Lori, and Jeremy Magruder. 2012. "Who Gets the Job Referral? Evidence from a Social Networks Experiment." *A.E.R.* 102 (7): 3574–93.

Bentolila, Samuel, Claudio Michelacci, and Javier Suarez. 2010. "Social Contacts and Occupational Choice." *Economica* 77 (305): 20–45.

Bewley, Truman F. 1999. *Why Wages Don't Fall during a Recession.* Cambridge, MA: Harvard Univ. Press.

Blau, Gary. 1990. "Exploring the Mediating Mechanisms Affecting the Relationship of Recruitment Source to Employee Performance." *J. Vocational Behavior* 37 (3): 303–20.

Brown, Meta, Elizabeth Setren, and Giorgio Topa. 2012. “Do Informal Referrals Lead to Better Matches? Evidence from a Firm’s Employee Referral System.” Staff Report no. 568, Fed. Reserve Bank New York.

Bryan, Gharad T., Dean Karlan, and Jonathan Zinman. 2015. “Referrals: Peer Screening and Enforcement in a Consumer Credit Field Experiment.” *American Econ. J.: Microeconomics* 7 (3): 174–204.

Burks, Stephen, Bo Cowgill, Mitchell Hoffman, and Michael Housman. 2015. “The Value of Hiring through Referrals.” *Q.J.E.* 130 (2): 805–39.

Calvo-Armengol, Antoni, and Matthew O. Jackson. 2004. “The Effects of Social Networks on Employment and Inequality.” *A.E.R.* 94 (3): 426–54.

Castilla, Emilio J. 2005. “Social Networks and Employee Performance in a Call Center.” *American J. Sociology* 110 (5): 1243–83.

Costa, Dora L., and Matthew E. Kahn. 2003. “Cowards and Heroes: Group Loyalty in the American Civil War.” *Q.J.E.* 118 (2): 519–48.

Dhillon, Amrita, Vegard Iversen, and Gaute Torsvik. 2012. “Employee Referral, Social Proximity, and Worker Discipline.” Working Paper no. 90, Univ. Warwick.

Dustmann, Christian, Albrecht Glitz, and Uta Schönberg. 2011. “Referral-Based Job Search Networks.” IZA Discussion Paper no. 5777, Inst. Study Labor, Bonn.

Fernandez, Roberto M., and Nancy Weinberg. 1997. “Sifting and Sorting: Personal Contacts and Hiring in a Retail Bank.” *American Sociological Rev.* 62 (6): 883–902.

Ghani, Ejaz, William R. Kerr, and Christopher Stanton. 2014. “Diasporas and Outsourcing: Evidence from oDesk and India.” *Management Sci.* 60 (7): 1677–97.

Gilchrist, Duncan, Michael Luca, and Deepak Malhotra. 2016. “When 3 + 1 > 4: Gift Structure and Reciprocity in the Field.” *Management Sci.* 62 (9): 2639–50.

Granovetter, Mark. 1995. *Getting a Job: A Study of Contacts and Careers.* Chicago: Univ. Chicago Press.

Heath, Rachel. 2015. “Why Do Firms Hire Using Referrals? Evidence from Bangladeshi Garment Factories.” Working paper, Univ. Washington.

Holzer, Harry J. 1987. “Hiring Procedures in the Firm: Their Economic Determinants and Outcomes.” Working Paper no. 2185, NBER, Cambridge, MA.

———. 1999. “Will Employers Hire Welfare Recipients? Recent Survey Evidence from Michigan.” *J. Policy Analysis and Management* 18 (3): 449–72.

Horton, John J. 2014. “Price Floors and Employer Preferences: Evidence from a Minimum Wage Experiment.” Working paper, New York Univ.

———. 2017. “The Effects of Algorithmic Labor Market Recommendations: Evidence from a Field Experiment.” *J. Labor Econ.*, forthcoming.

Ioannides, Yannis M., and Linda Datcher Loury. 2004. “Job Information Networks, Neighborhood Effects, and Inequality.” *J. Econ. Literature* 42 (4): 1056–93.

Kramarz, Francis, and Oskar Nordstrom Skans. 2014. “When Strong Ties Are Strong: Networks and Youth Labour Market Entry.” *Rev. Econ. Studies* 81 (3): 1164–1200.

Kugler, Adriana D. 2003. “Employee Referrals and Efficiency Wages.” *Labour Econ.* 10 (5): 531–56.

Lyons, Elizabeth. Forthcoming. “Team Production in International Labor Markets: Experimental Evidence from the Field.” *American Econ. J.: Appl. Econ.*

Montgomery, James D. 1991. “Social Networks and Labor-Market Outcomes: Toward an Economic Analysis.” *A.E.R.* 81 (5): 1408–18.

oDesk Corp. 2013. “Slide Deck with Statistics on the oDesk Marketplace.” http://www.marketwired.com/press-release/odesk-reaches-1-billion-spent-cumulatively-via-its-online-workplace-1817652.htm.

Pallais, Amanda. 2014. "Inefficient Hiring in Entry-Level Labor Markets." *A.E.R.* 104 (11): 3565–99.

Petersen, Trond, Ishak Saporta, and Marc-David L. Seidel. 2000. "Offering a Job: Meritocracy and Social Networks." *American J. Sociology* 106 (3): 763–816.

Pistaferri, Luigi. 1999. "Informal Networks in the Italian Labor Market." *Giornale degli Economisti e Annali di Economia* 58 (112): 355–75.

Regenstein, Marsha, Jack A. Meyer, and Jennifer Dickemper Hicks. 1999. "Job Prospects for Welfare Recipients: Employers Speak Out." Ser. A, no. 25, Urban Inst., Washington, DC.

Simon, Curtis J., and John T. Warner. 1992. "Matchmaker, Matchmaker: The Effect of Old Boy Networks on Job Match Quality, Earnings, and Tenure." *J. Labor Econ.* 10 (3): 306–30.

Stanton, Christopher, and Catherine Thomas. 2014. "Landing the First Job: The Value of Intermediaries in Online Hiring." Discussion Paper no. 1316, Centre Econ. Performance, London.

Topa, Giorgio. 2011. "Labor Markets and Referrals." In *Handbook of Social Economics*, edited by Jess Benhabib, Alberto Bisin, and Matthew O. Jackson, 1193–1221. Amsterdam: Elsevier.

Wang, Shing-Yi. 2013. "Marriage Networks, Nepotism, and Labor Market Outcomes in China." *American Econ. J.: Appl. Econ.* 5 (3): 91–112.