Why the Referential Treatment? Evidence from Field Experiments on Referrals

Amanda Pallais and Emily Glassberg Sands*
Harvard University and NBER; Harvard University

February 14, 2014

Abstract

This paper presents the results of three field experiments in an online labor market designed to determine whether referred workers perform better than non-referred workers and, if so, why. We hired workers and asked them to refer others; we then hired all referred and non-referred applicants. Referred workers performed better and had less turnover than non-referred workers. We find that this is partially due to selection: referrals outperformed non-referred workers even without any on-the-job interactions with their referrers. We find also that team production is important: referrals performed particularly well when working with their referrers.

*We would like to thank Kehinde Ajai, David Autor, Felipe Barrera-Osorio, 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, and seminar participants at Berkeley, Booth, Kellogg, Harvard, the New York Federal Reserve, Princeton, RAND, University of British Columbia, University of Chicago, Wharton, and NBER Summer Institute 2013 Labor Studies for their many helpful comments and suggestions. We would like to thank John Horton and the oDesk Corporation for help running the experiment. Sophie Wang provided excellent research assistance. Financial support from the Lab for Economic Applications and Policy at Harvard is gratefully acknowledged. Amanda Pallais's email address is apallais@fas.harvard.edu and Emily Glassberg Sands's email address is esands@fas.harvard.edu.

1 Introduction

A large empirical literature has shown that many workers find jobs through networks (e.g., Bewley, 1999; Ioannides and Datcher Loury, 2004; Granovetter, 1995). A consensus estimate is that at least half of jobs are found through informal contacts (Topa, 2011). Theoretical literature (e.g., Calvo-Armengol and Jackson, 2004; Montgomery, 1991) suggests that the use of referrals may disadvantage workers without labor market connections; consistent with this, empirical findings show that applicants who are not referred by current employees are much less likely than referred applicants to receive an offer (e.g., Fernandez and Weinberg, 1997; Peterson et al., 2000; Brown et al., 2012; Burks et al., 2013).

Yet the prevalence of referrals suggests that firms likely benefit from their use. Existing empirical work finds that referred workers have less turnover than non-referred workers (e.g., Brown et al., 2012; Holzer, 1987; Simon and Warner, 1992; Datcher, 1983; Burks et al. 2013). It remains divided, however, on whether referred workers are more productive. A few studies directly compare the performance of referred and non-referred workers working at the same or very similar firms: Castilla (2005) finds that referred workers perform better, Blau (1990) finds that they perform worse, and Burks et al. (2013) finds that referred workers perform better, but only on a few metrics. Other papers use wages or promotion rates to proxy for the performance of referred and non-referred workers; their findings are similarly mixed (e.g., Dustmann et al., 2011; Simon and Warner, 1992; Brown et al., 2012; Pistaferri, 1999; Bentolila et al., 2010).

We undertook three field experiments in an online labor market to identify whether referred workers perform better and have lower turnover than non-referred workers and, if so, why.[1] Our experimental approach affords us a unique opportunity to compare the performance of referred and non-referred workers without the filter of firms' hiring decisions. Most of the existing empirical literature compares the performance of referred and non-referred *hires* (where hires are a subset of all applicants). Differential selection of referred and non-referred workers into employment, however, complicates the interpretation of these results. For example, suppose a firm knows that referred applicants are on average more productive than non-referred applicants; a rational firm would incorporate this information into its hiring decisions such that, in order to be hired, a non-referred applicant would have to look relatively better on other characteristics. Indeed, the existing literature finds that even conditional on resume quality, referred workers are more likely than non-referred workers to

[1] There are other reasons firms might benefit from hiring referrals. For example, referrals might decrease the cost of recruiting or be a perk to existing (referring) workers. In this paper, we focus on productivity and turnover differences between referred and non-referred workers and abstract away from other potential benefits of hiring referrals.

be hired (e.g., Fernandez and Weinberg, 1997; Burks et al., 2013). Amid differential hiring by firms, hired referred workers may not perform any better than hired non-referred workers even when referrals provide positive information about worker quality.

In our experiments, we hired workers directly so that no differential employer selection could confound our comparisons between referred and non-referred workers. To recruit our experimental sample 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 non-referred workers to apply, and hired all applicants who met our basic wage criteria. Our design thus facilitates comparisons between referred and non-referred *applicants*.

In all three experiments (the "individual," "supplemental," and "team" experiments), we find that referred workers performed better than non-referred workers. Referred workers also had less turnover. These facts hold even conditional on resume characteristics; that is, referrals contained information about worker quality that was not contained in workers' resumes. The heart of this paper (and the motivation behind the three experiments) lies in assessing three potential explanations for these performance and turnover differences.[2] The first explanation, *selection*, says that a referred worker would perform better and stay longer at the firm even if she had not been referred. This may be, for example, because high-ability referrers also have high-ability friends (e.g., Montgomery, 1991; Granovetter, 1995; Rees, 1966), or because workers have information about their friends and select relatively productive and persistent contacts to refer (e.g., Beaman and Magruder, 2012; Fernandez et al., 2000; Rees, 1996).[3]

The second and third explanations, in contrast, emphasize how the productivity and turnover of referred workers may be affected by on-the-job interactions with their referrers. In the second explanation, *peer influence*, a referred worker exerts more effort and stays at the firm longer because she believes her performance and persistence will affect her friend's employment outcomes and/or their relationship. Consistent with this explanation, Heath's (2013) model suggests that referred workers work hard because if they perform poorly the firm will punish their referrers through lower wages; and Kugler's (2003) model assumes

[2]There is little empirical evidence on the mechanisms underlying performance and turnover differences between referred and non-referred workers. A prior version of Burks et al. (2013) entitled "The Value of Hiring through Referrals" analyzed potential mechanisms. But this discussion has been mostly removed from the current version, which focuses on observed differences (e.g., in offer rates, turnover, and performance) between referred and non-referred workers rather than on the mechanisms underlying these differences.

[3]Our experiments were designed so that referring workers had no information about the job itself at the time they submitted their referrals. Our results thus speak to selection on general observable and unobservable characteristics, and not to selection on match quality (i.e., workers referring friends who would be a good fit for the particular job).

referrers directly exert peer pressure on their referrals to perform well.[4]

In the third explanation, *team production*, a referred worker performs better and may enjoy her job more when working directly with her referrer. While this explanation for referrals' positive performance has not been emphasized to the same extent in the economics literature, general research on team production implies it may be an important benefit of referrals. Bandiera et al.'s (2012) model, for example, finds that when working in teams with their friends, workers receive more utility and are less likely to free-ride. Furthermore, Bandiera et al. (2005) finds that workers are more able to cooperate with their teammates when their teammates are friends; and Costa and Kahn (2003, 2010) find that Civil War soldiers were less likely to desert and were more resilient to job-related stress when more of their unit was from their own birthplace.

Our three experiments are designed to test these three explanations: the individual experiment distinguishes between selection and peer influence, the supplemental experiment explores selection more deeply, and the team experiment isolates team production. We find that selection is important. On-the-job interactions between referrers and their referrals are also important; while we see only limited evidence of peer influence on the job, we find substantial evidence of team production.

The individual experiment distinguishes between selection and peer influence. All referred and non-referred workers in this experiment performed an individual task: testing an airline flight website by answering a few questions about the flights listed on the site every other day over the course of 12 days.[5] Referrers were simultaneously completing a different task and were randomized, along with their referrals, into one of two treatment groups. Treatment 1 was designed to maximize peer influence. For example, each referrer in this treatment received an update on her referral's performance after each day of work and the referred worker knew her referrer was receiving these updates. We implied to each referrer that her referral's performance and willingness to continue working for us would affect whether she was promoted. Treatment 2, in contrast, was designed to minimize peer influence. Each referred worker in this treatment was told her referrer would never know how she performed and referrers were told explicitly that they would be judged on their own merits, *not* on the performance of their referrals. At the end of the job, we asked each worker if she would like to continue with the firm.

From the individual experiment, we learn that selection is important. Even non-monitored (Treatment 2) referred workers performed better and stayed longer than non-referred work-

[4]The peer influence explanation is also related to group lending in microfinance wherein a worker's peers may pressure the worker to repay the loan (e.g., Bryan et al., 2012).

[5]The tasks for all three experiments were chosen to be similar to tasks that are common on oDesk. In particular, many jobs on oDesk require visiting websites and answering questions about them.

ers. We also find that the referral provides information to employers that could not easily be obtained through observables or initial job performance: the non-monitored referred workers had better observable characteristics than non-referred workers, but they outperformed and outlasted their non-referred counterparts conditional on these.[6] They even outperformed non-referred workers on the last day of the contract, controlling for their performance on all of the prior days. Comparing Treatment 1 (monitored) and Treatment 2 (non-monitored) referred workers, we do not find evidence that peer influence had large effects on workers' productivity or persistence: monitored referred workers performed slightly better, but the difference was not statistically significant. They were, if anything, slightly *less* likely to want to continue working for the firm, perhaps because they disliked being monitored.

Since even referred workers who were not monitored may have faced some subtle peer influence in the individual experiment, we ran a supplemental experiment four months later to isolate the effects of selection. We made job offers to all referred and non-referred workers from a new firm that had no affiliation with the firm from the individual experiment and had no contact with any of the referrers. The task was designed to be credibly different from that in the individual experiment, though it similarly measured diligence over time and willingness to stay on at the firm.

The supplemental experiment provides the strongest evidence that selection is a key driver in the superior performance and persistence of referred workers. Even at a firm to which they had not been referred and at which their referrers did not work, referred workers exhibited substantially higher performance and lower turnover than non-referred workers. The effects are generally large and significant regardless of whether we restrict attention to workers who accepted our offer of employment.

Our third experiment, the team experiment, isolates the effect of team production. 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 written in all capital letters, be exactly three words long). We asked the partners to use the chat box provided on the site to discuss the task and then to each submit the same slogan, which should have satisfied both criteria. Workers completed three such PSA tasks, each with a different partner. Every referrer participated exactly once in each of three team types: a Type A team (where she was paired with her own referral), a Type B team (where she was paired with someone else's referral), and a Type C team (where she was paired with a non-referred worker). We

[6] The online marketplace in which these experiments take place is a unique setting in that we see workers' entire resumes. Because interviews are relatively uncommon and workers and employers do not meet face-to-face, we observe most characteristics that an actual employer would observe when making its hiring decisions.

measured performance in each pairing and, after all three PSA tasks had been completed, also asked with which partner(s) they would want to work again.

We find substantial evidence of team production. Referred workers outperformed non-referred workers even when both types were assigned partners they did not previously know, but referred workers performed substantially better still when paired with their own referrers. They also worked longer on the task when paired with their own referrers and were more likely to report wanting to continue working with their own referrers than with their other partners. These results suggest team production is an important benefit of hiring referrals.

Across experiments we find that referrals provide (positive) information about worker performance on top of workers' observable characteristics, but not all referrals are created equal. Workers referred by high-performers performed particularly well themselves. Part of this can be explained by a tendency among referrers to refer workers with observable characteristics similar to their own: referrers with stronger resumes on average provide referrals with stronger resumes. But even controlling for workers' observable characteristics, those referred by high-performers tended to perform better themselves. We also explore the relationship between the strength of the referrer-referral tie and the performance of the referred worker. At the time of referral, we asked the referrer three questions about her relationship with her referral: how well she knows her referral, how many friends they have in common, and how often they interact. (A caveat is that these were self-reported before the referral had been hired.) We find that when a worker refers someone with whom she is not as close (a weak tie), she tends to refer someone who looks better on paper. Nevertheless, it is the referral who has a stronger tie to her referrer who performs better, even before controlling for observable characteristics.

Finally, we use our experimental data to show that if we had only compared the performance of those referred and non-referred applicants whom employers had actually chosen to hire, we could have obtained misleading results about the information contained in a referral. We first simulate which of our applicants employers would choose to hire if they observed both resumes and referral status, assuming they knew the relationship between resumes and referral status, and performance. Because referred workers substantially outperformed non-referred workers conditional on observable characteristics, employers would hire relatively few non-referred workers and the non-referred workers hired would be very positively selected on observables. We then compare the actual performance of the workers hired in our simulations. We show that even though the referral contained important information about worker quality, there would be no significant difference in the performance of *hired* referred and non-referred workers.

All three field experiments took place on oDesk, the largest online labor market, with over

2.5 million workers (Horton, 2013) and 35 million hours billed in 2012 (oDesk Corporation, 2013). In this context, we were able to hire workers directly, thus eliminating the concern that employers differentially selected referred and non-referred workers into employment. Equally important, the online labor market allowed us to carefully alter the parameters of the jobs and what workers observed in order to tease out the effects of selection, peer influence, and team production in ways that would be very difficult to execute effectively in brick-and-mortar firms. The trade-off is that our results come from a specific labor market. Before detailing the experiments or their results, we first describe the marketplace (Section 2). We also discuss external validity and the main way we think oDesk differs from more traditional labor markets: oDesk workers are often less strongly tied to employers than are workers in other labor markets. Selection and peer influence may thus be less important on oDesk than in other markets. For example, if oDesk workers are less concerned with remaining in good standing with their employers, they may not refer particularly talented workers (selection) or put pressure on their referrals to work hard (peer influence). Thus, given that we find that selection is important even on oDesk, it seems likely that selection is also quite important in other contexts. However, the fact that we don't find strong evidence of peer influence on oDesk does not eliminate the possibility that it is important in other contexts.

After describing the marketplace and discussing external validity, Section 2 also explains the sample selection for our experiments and provides descriptive statistics about our sample. The three subsequent sections describe the design and results from the individual experiment (Section 3), the supplemental experiment (Section 4), and the team experiment (Section 5). Section 6 analyzes how referrers' performance and the referrer-referral relationship predict referred workers' performance. Section 7 shows that the comparison of referred and non-referred workers' performance could be biased if we only observed the performance of workers employers chose to hire and Section 8 concludes.

2 Experimental Context and Recruitment Design

2.1 Online Labor Market

oDesk is an online labor market where employers, mostly from the United States, hire independent contractors from all over the world for tasks ranging from software development to administrative support.

Employers post job listings and can invite workers to apply; workers, meantime, post online resumes and bid on those jobs. Resumes typically include previous oDesk jobs, a one-

to-five feedback score from these jobs, and an hourly wage suggested by the worker. Many also list other qualifications such as degrees held and oDesk tests passed. Figure 1 shows a sample oDesk resume. (This worker was not in our experiment.)

Employers decide which of the two oDesk job types they want to offer: hourly jobs or fixed-wage jobs. Hourly jobs are the more popular type of job and the type used throughout our experiments. In these jobs, workers propose an hourly wage when they apply. Employers choose which workers to hire. Hired workers track the time they are working and oDesk monitors that they are actually working during these periods by taking screenshots and analyzing keystroke volume. Workers are then paid their set hourly wage for the time worked regardless of output quality.[7]

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.

2.2 External Validity

Completing these experiments in an online labor market allows us to observe the performance and persistence of workers without the filter of firms' hiring decisions. It also allows us to vary parameters of the jobs workers completed to cleanly identify why referred workers perform better and have less turnover than non-referred workers. The trade-off, however, is that the results of this experiment come from one particular labor market. Perhaps the biggest difference between oDesk and other labor markets is that because oDesk jobs are relatively short and oDesk workers work for many employers, oDesk workers are less tied to any particular employer than are workers in other markets. Prior to our experiment, the average job taken on by the referrers in our sample paid $237 and lasted 81 working hours.

The fact that oDesk workers are less tied to any particular employer could mean that selection and peer influence are weaker here than in other markets. For example, suppose that selection stems from workers choosing their most talented friends to refer (as opposed to homophily in friend networks). If workers are not as tied to employers, they may be less careful to refer only their particularly talented friends. Similarly, if referrers are not as worried about their standing with the employer, they may exert less pressure on their referrals to perform well.

Because we were concerned that peer influence might not be as strong a motivator on oDesk as in other labor markets, we aimed to maximize the effect of peer influence in Treatment 1 of the individual experiment. It is hard to imagine another context wherein a worker's

[7] In contrast, in fixed wage jobs workers and employers agree on a price for the entire job and employers have discretion at the job's end over how much to actually pay.

promotion would be so closely tied to her referral's performance. Still, in this experiment, we find limited effects of peer influence. Despite this, our findings do not rule out the fact that peer influence may be important in other markets.

However, the fact that selection might be less important in oDesk than in other labor markets is of limited concern. Given that we find strong evidence that selection is important on oDesk, it seems likely that it is important in other markets as well.

2.3 Hiring our Experimental Samples

We hired workers for the individual and team experiments in the same way. 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 resume, (3) had earned $50 or more on oDesk, and (4) had an average job feedback score of four or higher to apply to our job. We eliminated workers with ratings below four because we only wanted 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 four.[8] 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 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.

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 whom they thought would "do a good job and be interested in an ongoing relationship with our firm." On each referral form we included questions about how well the referrer knew her referral, how often they interacted (remotely and/or 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 and/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

[8] We only included workers from the Philippines because we wanted all workers in the team task to be able to communicate easily and the Philippines is the most common country of residence for low-wage oDesk workers doing these types of jobs.

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.[9] We again gave workers 48 hours to apply and then hired all referred and non-referred workers who applied at an hourly wage of $3 or less.[10]

This recruiting process, used for both the individual and team experiments, produced an experimental sample with three types of workers: referred workers, non-referred workers, and "referrers" (i.e., workers who made a successful referral). 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 supplemental experiment, we made job offers to all referred and non-referred workers from the individual experiment; no referrers were included.

2.4 Descriptive Statistics

Table 1 describes the characteristics of three groups of workers: (1) all referred workers, regardless of whether they met our criteria, (2) included referred workers (i.e., referred workers who met our criteria and applied at a wage of $3 or less) and (3) included non-referred workers (i.e., non-referred workers who met our criteria and applied at a wage of $3 of less).[11] Included referred workers had, on average, been on oDesk for about 15 months. Almost two thirds had prior oDesk employment; those who had been employed, had, on average, about nine jobs and earned about $1,350. Non-referred workers had been on oDesk for about four months longer, but were much less likely to have been previously hired; only

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

[10] We designed the recruitment process so that when referrers were submitting their referrals, they had no information about our actual tasks. The initialization step, for example, 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 non-referred workers saw this same description on our job posting. Since referred workers had no private information about the job before referring, in our context there is no scope for selection on match quality.

[11] While we hired all referred workers who met our criteria and applied at a wage of $3 or less, only one (randomly-selected) referred worker per referrer was actually included in the team experiment. (The remaining referred workers completed the same tasks, but with different partners. Their performance data is not presented.) Thus, there are some referred workers who applied and met our hiring criteria but are not considered "included referred workers." After all three experiments, we were required by the Harvard IRB to inform all participants about the study and give them the opportunity to remove their data from our study. One worker who was referred but had been excluded from our experiments requested to have his data removed and we have done so. Removing this worker's data only affected the "All Referred Workers" column in Table 1 and the "Excluded Referred Workers" column in Appendix Table 8.

28 percent had prior experience. Referred workers also appeared to be more qualified than non-referred workers: they had higher feedback scores from prior employers, had passed more oDesk tests, and had higher self-assessed English abilities. Despite being seemingly more experienced and qualified than non-referred workers, referred workers posted wages on their resumes that were over 20 percent *lower* than those posted by non-referred workers, and they proposed significantly lower wages to our jobs. Referred workers were also much more likely to apply to our job: 68 percent of referred workers applied versus only six percent non-referred workers. (The six percent of non-referred workers who took the time to apply were themselves a very positively self-selected group.) This suggests referrals are a way to identify workers with good resumes who are interested in the job.[12]

3 Individual Experiment

3.1 Design: Identifying Selection and Peer Influence

The task for the individual experiment was designed to measure referred and non-referred workers' diligence when working alone on a project. We designed our task 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 markets and on oDesk.[13] 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., Brown et al., 2012; Holzer, 1987; Datcher, 1983). The treatments were designed to determine the extent to which observed differences in workers' performance and turnover were driven by selection relative to peer influence.

All referred and non-referred workers completed the same task. We told them they would be doing testing for an airline flights website, and asked that they visit the site every other day for twelve 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 ten 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

[12] Appendix 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.

[13] For example, on oDesk, Pallais (2014) finds that employers care more about whether a worker completed a data entry task by the deadline than the worker's accuracy. In more general labor markets, firms respond to absenteeism by having other employees work overtime, reassigning workers from other jobs, and/or hiring temporary workers. These adjustments are all costly and often require manager time. Moreover, the replacement workers may not be as productive as the absent workers (e.g., Herrmann and Rockoff, 2010).

(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. Appendix Figure 1 displays a sample flights table followed by the questionnaire.

We told all referred and non-referred workers to complete the task on the assigned day and asked, but did not require, that they complete each day's task by 11am Philippine Time. We also informed all referred and non-referred 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 11am 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. Appendix Figure 2 shows an example performance report.

Referrers were randomized to Treatments 1 and 2. Each referred worker was assigned the same treatment as her referrer. Appendix Table 2 shows that the randomization produced balanced samples between Treatment 1 and Treatment 2 within both the referrer and referred worker samples. Out of 26 comparisons between the two treatments groups, only one difference is significant at the 10 percent level.[14]

Treatment 1 was designed to facilitate monitoring of the referred worker by her referrer while Treatment 2 was designed to minimize peer influence on the referred worker. Referred workers in Treatment 1 were told that their daily performance statistics would be sent to their referrer as well as the manager. Referred workers in Treatment 2, meantime, were explicitly told that their referrer would never see their performance statistics, only the manager would.

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 would be sending them about their referrals. Thus, we asked them to answer questions on a website every other day over the same time twelve day period and we assigned them a soft deadline for submitting on each day. 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 login method, was focused on consumer products rather than flights, and asked a different set of questions; referrers also had a different soft deadline (2pm Philippine Time).

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 Treatment 1 that whether they were promoted would depend on their referrals' performance.[15] Referrers

[14] 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.

[15] 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 Treatment 1 about the position, we also said that they would receive daily performance updates on their referrals "because we

in Treatment 2 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 Treatment 1 to her respective referrer. We also sent referred and non-referred workers' statistics to a manager we hired.

At the end of the task, we invited all referred and non-referred workers to re-apply to continue on the same project. We use this as an (inverse) measure of worker turnover. Re-application updates by worker type and treatment mirrored performance updates. Each referred and non-referred worker was told that the manager would receive an update on whether she accepted our offer to re-apply. Referred workers in Treatment 1 were told this update would also go to their referrers while referred workers in Treatment 2 were explicitly told their referrers would not see this information. To strengthen the treatment, when we invited referrers in Treatment 1 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 Treatment 2 to apply for the management position as well, but made no mention at all of their referrals.

3.2 Performance and Persistence by Worker Type and Treatment

Table 2 shows how monitored referred workers (Treatment 1), non-monitored referred workers (Treatment 2), and non-referred workers compare on three measures of performance and our measure of persistence with the firm (the inverse of turnover).[16] Each column presents the results of regressing an outcome on an indicator for being a referred worker in Treatment 1 (a referred worker monitored by her referrer), an indicator for being a non-referred worker, and workers' observable characteristics.[17] The omitted group is referred workers in Treatment 2 (non-monitored referred workers).

Although referred workers had more positive observable characteristics than non-referred workers, the referral still had substantial predictive power even conditional on these char-

care about workers' performance."

[16] Two of the performance metrics 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 11am. Workers were also told that the manager would see whether the worker answered all questions, but we exclude this metric from our analysis since 99.8 percent of submissions were complete. The final performance metric is accuracy (non-responses are marked as incorrect).

[17] The observable characteristics included in the regressions are as follows: 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 a portfolio, the number of oDesk tests passed, 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.

acteristics. Referred workers consistently outperformed non-referred workers even when the referred workers were not monitored. For example, non-monitored (Treatment 2) referred workers submitted responses on just over three-quarters of assigned days; conditional on observable characteristics, non-referred workers were 13 percentage points less likely to submit. Referred workers were also much more likely to want to continue with our firm. While almost all referred workers in Treatment 2 (95 percent) wanted to continue working with us, non-referred workers were 22 percentage points less likely to re-apply to continue the task (conditional on observables).

Across the three performance metrics, the coefficients on the Treatment 1 dummy suggest that peer influence may have led referred workers to perform better still. Anecdotal evidence suggests that referred workers in Treatment 1 were, in fact, monitored by their referrers. Many Treatment 1 referrers 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 and/or had not completed it by the soft deadline, and assured us they would encourage their referrals to do better on subsequent days. Nonetheless, all of the performance differences between referred workers in Treatments 1 and 2 appear smaller than the differences between the referred workers in Treatment 2 and non-referred workers, and none is significant. The negative (though again insignificant) coefficient on the Treatment 1 dummy in the final column suggests that referred workers in Treatment 1 were, if anything, slightly less likely to be interested in continuing with the firm, perhaps because they disliked being monitored.

Throughout the paper, we use the covariates listed in footnote 17 as our main controls. Our results are, however, robust to adding the squares of each of the (non-binary) covariates and the interaction of each pair of covariates to the regressions (what we call "second order controls"). The first two panels of Appendix Table 3 shows the results of replicating Panel A of Table 2, eliminating all the worker controls (Panel A) and adding the second order controls (Panel B). When the second order controls are added, two of the coefficients on the non-referred dummy increase and two decrease, though none changes significantly. Unsurprisingly given random assignment, adding control variables does not affect the estimated differences between monitored and non-monitored referred workers.

Referrals provide the firm with information about worker quality; firms might also get information about a worker's quality by hiring the worker for some trial duration and observing her performance directly. Longer trials almost certainly provide better information, but at a direct wage cost to the firm. Panel B 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 B replicates Panel A, limiting the observations to the last day of

the contract. Regressions in the first three columns now additionally control for the worker's performance (on the same metric as measured by the dependent variable) on each of the first five days. All differences in performance between referred and non-referred 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. The final column of Panel B adds controls for each of our performance measures (submission, on-time submission, and accuracy) on each of the six days. Even controlling for all our performance measures on all days, referred workers were 18 percentage points more likely than non-referred workers to want to continue on with the firm.[18] Panel C of Appendix Table 3 shows that these results are robust to adding the second order controls.

Figure 2 shows worker performance over the course of the experiment by worker type and treatment. Submission rates of referred workers were consistently higher than those of non-referred workers. Both types of workers became less diligent over time, but diligence fell off much more for non-referred workers. Thus, the performance gap between referred and non-referred workers grew with time. The performance gap between Treatment 1 and Treatment 2 referred workers was less stark. On the first day of work, before any performance reports had been sent out, monitored and non-monitored referred workers performed equivalently. The graph suggests that peer influence may have stemmed the drop-off in performance in days two, three, and four among Treatment 1 referred workers, though the differences between monitored and non-monitored referred workers on these days is not significant. By day six, however, monitored referred workers were no more likely than their non-monitored counterparts to submit.

Taken together, the results of the individual experiment suggest that selection is important. Even when referred workers were not monitored by their referrer, they performed much better than non-referred workers and were more eager to continue on with the firm. The referral, moreover, contained information that was not present on a worker's resume or in her performance on the majority of her contract. In contrast, 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.

[18] Unreported coefficients in the final column of Panel C show that workers who performed better were more likely to want to continue.

4 Supplemental Experiment

4.1 Design: Isolating Selection

Even though all referrers in Treatment 2 of the individual experiment were assured that they would be judged only on their own merits and all referred workers in this treatment were assured that their referrers would not see their performance statistics, these referred workers may still have been influenced by the presence of their referrers at the firm. They may have, for example, felt grateful for having been referred or faced informal pressures from their referrers, either of which could have affected their performance or persistence in the individual experiment. The supplemental experiment was designed to eliminate any such potential influences.

In the supplemental experiment, we measured the performance and persistence of referred and non-referred workers in a job to which the "referred workers" had not been referred. Four months after the individual experiment, we created a firm with a different name, location, job posting, and writing style from that of the individual experiment. None of the referrers was contacted by this firm. To minimize selection, we sought to hire the maximum possible number of referred and non-referred workers. We made direct job offers to all referred and non-referred workers from the individual experiment and sent three reminders to accept to workers who had not yet responded.

Workers who accepted were given a task that, like the task of the individual experiment, measured individual diligence over time. They were asked to visit the Twitter pages of three successful musicians and to answer a ten-question survey about those accounts every day for five 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. Appendix Figure 3 displays the site and questionnaire. After the last assigned day of work, we again invited workers to a continuation of the task and recorded whether they re-applied.

4.2 Pure Selection Effects

The majority (61 percent) of workers from the individual experiment accepted our offer and so were hired for the supplemental experiment; referred workers were significantly more likely to accept than non-referred workers. However, regardless of whether we include all referred and non-referred workers in the analysis (counting as not submitting work those who did not

accept our employment offer) or instead analyze performance conditional on accepting our job offer, our key results remain unchanged: even working at a job for which they were not referred at a firm with which their referrers were not affiliated, referred workers outperformed non-referred workers and had less turnover.

The three performance metrics in the supplemental experiment mirror those of the individual experiment: an indicator for submitting a response for a given day, an indicator for submitting that day's response on the correct day (analogous to the soft deadline of the individual experiment in that it was requested, but not required), and the fraction of questions answered correctly. Panel A of Table 3 shows that unconditional on accepting our employment offer, referred workers were 9 percentage points more likely to submit a response and to submit it on the correct day, even conditional on their observable characteristics. In addition to performing better, referred workers were 12 percentage points more likely than non-referred workers to apply for a continuation of the task.

Next, we compare the performance of referred and non-referred workers who accepted our job offer. Appendix Table 4 provides suggestive evidence of, if anything, differentially positive selection of non-referred workers into accepting. In this table, we regress an indicator for accepting our job offer in the supplemental experiment on an indicator for being non-referred, a metric of performance or persistence in the individual experiment, and the interaction of that metric and the non-referred dummy. Each column uses a different performance or persistence metric from the individual experiment. The large standard errors render many of the results statistically insignificant, but the coefficients suggest that there was positive selection of non-referred workers relative to referred workers into accepting our job.

Panel B of Table 3 shows the results of estimating the same regressions as in the table's Panel A, now limited to workers who accepted our job offer. In this sample, referred workers were 10 percentage points more likely to submit work and 13 percentage points more likely to re-apply. However, given the results in Appendix Table 4, we might expect the conditional results to slightly underestimate the true performance and turnover differences between referred and non-referred workers.

The magnitudes of these estimates are similar to those from the individual experiment. Referred workers performed about as well here as did the non-monitored (Treatment 2) referred workers in the individual experiment (both submitted 76% of days, for example). The performance gap between non-referred workers and referred workers is also similar to that of the individual experiment. Appendix Figure 4 shows that, as in the individual experiment, the gap between referred and non-referred workers widened over the contract, while Appendix Table 5 shows that the results in Table 3 are robust to the addition of the second order controls. It also shows results without controls for worker characteristics.

5 Team Experiment

5.1 Design: Identifying Team Production

The team experiment was designed to measure whether referred workers outperform non-referred workers in a task involving teamwork and, if so, to determine how much of the performance difference between referred and non-referred workers is due to the fact that referred workers may perform particularly well when working with their referrers (team production).

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 public service announcements (PSAs).[19] 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. Appendix Figure 5 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 justification was that there was a lot of information to process and that by giving the partners different information, each partner would only have to read 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 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 their 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, the other partner's word choices

[19] As with the prior tasks, we chose this task because there are many jobs on oDesk that ask low-skill workers to come up with advertisements, for example in the form of flyers, posters, and/or slogans.

emphasized that flu shots are effective. When giving workers their information sheets, we told them only that the sheets would contain information, not 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.[20]

For comparison with the individual and supplemental 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 seatbelts). 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 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 non-referred worker. 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.) Non-referred workers worked with referrers for all three rounds; that is, they were always in Type C teams.

Because we thought worker performance might be correlated not just between partners, but also among partners' partners, we placed workers into blocking groups. By definition, every worker in the blocking group only ever partnered with others in the same blocking group. In all analyses of the team experiment, we cluster standard errors by blocking group.

Each of the 47 blocking groups contained six referrers, their six referred workers, and two non-referred workers. The placement into blocking groups was random, except that a referrer and her referral were always in the same group.[21] Within a blocking group, the

[20] Because we wanted to measure how effectively workers worked with their partners, we strongly encouraged each worker to complete each PSA. Unlike in the individual 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.

[21] As in the individual experiment, we hired all referred and non-referred workers who met the selection

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.

5.2 Performance by Team Type

Panel A of Table 4 compares referred and non-referred worker performance across team types 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.[22]

Each outcome 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 non-referred worker paired with a referrer). Controls for the referred and non-referred worker's own characteristics are included throughout. The omitted group contains workers in Type B teams (referred workers paired with someone else's referrer). Thus, the coefficient on the Type A dummy indicates how much better referred workers perform when paired with their own referrer than with someone else's referrer; the Type C dummy indicates how much worse non-referred 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 non-referred workers. Referrers' performance does not vary significantly across team types.

On average, referred workers performed well on these diligence measures. When paired with someone else's referrer, referred workers logged in 88 percent of the time, submitted work 84 percent of the time, and correctly answered their own question 76 percent of the time. Less than half (44 percent), however, included the criteria from their own information sheet in their slogan.[23] Non-referred workers, meantime, were substantially less diligent than referred workers, even when neither group was working with a partner they previously knew.

criteria. However, only one randomly-selected referral from each referrer and only 94 non-referred workers were included in this experiment.

[22] 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 criteria in the slogan.

[23] For comparison, referrers on these teams were four percentage points more likely to log in and eight percentage points more likely to use their own criteria in their slogan.

As compared to referred workers in Type B teams, non-referred workers (all on Type C teams) were approximately 20 percentage points less likely to log in to the site, to submit a slogan, and to correctly answer their individual question, even conditional on their observable characteristics.

The coefficients on Type A teams show that referred workers were five percentage points more likely to submit their work and to correctly answer the question about their own reading when they were paired with their own referrer instead of with someone else's referrer. Given that these are measures of diligence more than teamwork, it could suggest that peer influence may have played a role in the team task. When working together, referrers may have put more pressure on their referrals to be diligent because in this context, their referrals' performance affected their own.

Panel B compares team performance by team type. Observations are again at the partner-pair level. It shows that, on measures of team performance, teams with a referred worker consistently outperformed those with a non-referred worker, even when the referred worker was working with someone else's referrer. For example, while half of Type B teams answered the team question (e.g., what color sign the PSA should be printed on) the same way, Type C teams were 17 percentage points less likely to do so.

While referred workers did well relative to non-referred workers even when not working with their referrers, they did particularly well when working with their referrers. Referred workers were, for example, 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 question. The results are consistent across team performance metrics. The third column shows similar results for submitting the same slogan. Only about one-third of Type B teams submitted the same slogan. Type C teams were about a third less likely to do so; Type A teams were more than twice as likely.[24] Appendix Table 6 replicates this table, both removing the individual controls and by adding the second order controls.

[24] One potential explanation for why referred workers performed better when working with their referrers is that a referred worker and her referrer were, on average, more similar than a randomly-selected referrer and referred worker. We find no evidence, however, that this drives our results. 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. But 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.

5.3 Enjoyment and Time Spent by Team Type

One potential motivation for hiring referrals is that workers might enjoy working with their friends and, thus, might be willing to spend more time on the job. Because oDesk requires workers to record the time they spend working on oDesk tasks, we can analyze the amount of time workers spent on each of the three PSAs. Panel A of Table 5 shows time spent on the task by team type, first for referrers and then for referred and non-referred workers.

When partnered with someone they did not know, referrers spent the same amount of time (around 37 minutes) on the task regardless of whether their partner was a non-referred worker or someone else's referral. When working with their own referral, however, they spent an average of six extra minutes on the task. Referred workers also spent significantly (14% more) time on the task when working with their referrers.

In general, workers who spent more time on the task performed better. Even controlling for the time workers spent on the task, however, Type A teams performed better than Type B teams. A separate but related reason Type A teams might have performed better is that they communicated via different methods. While each worker always had access to a chat box on the site in which she could chat live with and/or leave messages for her partner, Type A teams may have been advantaged by having other means of communicating. While Type A teams did communicate more both inside and outside of the chat box, this cannot explain their superior performance. Appendix A describes these analyses in more detail.

Panel B of Table 5 provides additional insight into how much workers enjoyed their work experience on each type of team. After they had completed all three tasks, workers reported which partner(s) they would be interested in partnering with again; workers could choose as many or as few partners as they wanted.[25] We find that referrers were significantly more likely to want to work again with referred workers they did not know than with non-referred workers.[26] But, referrers were more than twice as likely to want to partner again with their own referral as with someone else's referral. Similarly, referred workers were substantially more likely to want to work again with their own referrer than with someone else's referrer.

6 Predictors of Referral Performance

We find across our experiments that having been referred is a powerful, positive predictor of performance: in each of our three experiments, referred workers substantially outperformed

[25] Some workers (about 20 percent) did not answer the question, mostly because they did not complete the third PSA task. But for those who answered, we know whether or not they wanted to work again with each of their three partners.

[26] Referrers did not know who, besides their own referrals, had been referred to the firm.

their non-referred counterparts. But not all referrals are created equal. In this section we focus on referred workers and identify predictors of their performance. We look first at a referrer's performance as a predictor of the performance of her referral and then turn to the relationship between the referrer and her referral.

6.1 Referrer's Performance

The first column of Table 6 shows that a referrer's performance is a strong predictor of her referral's performance. We regress the referred worker's performance in the individual experiment on her referrer's performance in the same experiment. (We use submission as our performance metric here, but using other performance metrics provides similar results.)

What this result does not illuminate is *why* the performance of the referring worker is a good predictor of her referral's performance. For example, referrers and referred workers may perform similarly because on any given day they experience common shocks, or because they have similar underlying ability or diligence. In fact we find that (1) even absent common shocks, workers tend to refer people who perform as they do, (2) part of this seems to be driven by the positive correlation between a worker's own observables and those of the worker she refers, and (3) even absent common shocks and controlling for the referred worker's observable characteristics, the referrer's performance still predicts her referral's performance.

Because it was executed four months after the individual experiment, the supplemental experiment allows us to disentangle the common shocks hypothesis from others. (This assumes that common shocks do not persist for four months.) In the second column of Table 6, we regress the referred worker's performance in the supplemental experiment on her referrer's performance in the individual experiment four months earlier. Even absent common shocks, the referrer's performance remains a powerful predictor of the referral's performance. In fact, knowing the performance of a worker's referrer four months prior leads to almost two-thirds as much updating as knowing her own performance four months ago (Table 6).

Only some of this can be accounted for by observables. Appendix Table 7 shows that workers with better observable characteristics refer workers who also have better observables. Controlling for the referred worker's observables in the regression in Table 6 reduces the point estimate on the referrer's performance by about half. Nonetheless, the referrer's performance remains a large and positive (albeit not statistically significant) predictor of her referral's performance.[27] This suggests that higher performers refer workers who perform better than would even be expected based on their observable characteristics.

[27] When on-time submission is used as the performance metric instead of submission, this coefficient is significant at the five percent level.

6.2 Strength of Referrer-Referral Relationship

We turn now to the relationship between referrers and their referrals. Appendix Table 8 shows the distributions of the three relationship variables we have from referrers' reports at the time of the referral.[28] Referrers tended to refer workers they were close to. Among those included in the experiment, most reported knowing their referrals "extremely well" (six on a scale of one to six), while only one percent said they knew their referral "hardly at all" (one 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. We also asked workers how many other people they knew in common with their referral: 48 percent of referred workers knew 20 or more people in common with their referrer.

Because each relationship variable is consistently a positive predictor of the referral's performance, we build an index of relationship strength and for parsimony focus here on the resulting estimates.[29] We exclude the five referred workers whose referrers did not answer all the relationship questions at the time of the referral.

Panel A of Table 7 shows how characteristics of the referred worker vary with the strength of her relationship with her referrer. Each column shows the results of a different observable characteristic regressed on the relationship index. The reported coefficients show that referred workers who have stronger relationships with their referrer look worse on observables. They have passed fewer oDesk tests, completed fewer assignments, and, conditional on receiving feedback, have received (insignificantly) worse feedback.[30] These results suggest that when referrers refer people with whom they have weaker ties, they refer people who look better on paper.

Panel B investigates how a referral's performance varies with the strength of the referrer-

[28] We caveat this section by emphasizing that these relationship characteristics are self-reported by referrers. Referrers could have reported being close with workers whom they thought would perform particularly well and/or whom they particularly wanted the employer to hire.

[29] In building the index, we first create dummy variables for reportedly knowing the referred worker well (responding more than three on a scale of one to six when asked how well she knew the referred worker), interacting with the referral at least once a week, and knowing at least twenty people in common. Our relationship index is defined as the standardized sum of these three binary variables. The magnitudes of the coefficients are similar if we define the index instead as the standardized average of z-scores for the three raw variables (on scales of 1 to 6 for how well the referrer knew her referral, 1 to 7 for how often they interacted, and 1 to 5 for how many people they knew in common) or of z-scores for three constructed variables (with how well they knew each other on the same scale, but with how often they interacted coded as the estimated number of days per month they interacted and with how many people they knew in common coded as the midpoint in the chosen range).

[30] They are also significantly (5 percentage points) less likely to have received any feedback, probably because they have completed fewer assignments. Point estimates suggest they also look worse on the other observable characteristics we have, but the coefficients are not generally significant.

referral relationship. For each experiment, a worker's performance on a given day (or a given PSA in the team experiment) is regressed on the relationship index. For parsimony we present only one outcome per experiment, though within experiments, the magnitudes of the coefficients are similar when performance is defined using the other metrics. In each experiment, referred workers performed better the stronger their relationship with their referrer. A referred worker with a one standard deviation stronger relationship with her referrer was four percentage points more likely to submit work in the individual experiment and two percentage points more likely to submit work in the supplemental experiment, though the latter point estimate is not statistically significant. In the team experiment a referred worker with a one standard deviation stronger relationship with her referrer was five percentage points (nearly ten percent) more likely to have her slogan match her partner's.

Panel C presents the results of these same regressions with the inclusion of controls for the referred worker's observable characteristics. Given that referred workers with stronger ties to their referrers tended to have worse observable characteristics, it is unsurprising that the coefficients on the relationship index are, on average, larger when the controls are added.[31]

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 observables.

7 Potential Bias from Employers' Hiring Decisions

In each experiment, we hired all applicants who met our basic hiring criteria. This ensures that employers' hiring decisions did not lead to differential selection of referred and non-referred workers into our sample. In this section, we use our experimental data to simulate how our comparisons between referred and non-referred workers might have been biased had we only observed the performance of workers an employer chose to hire. Though the results of this exercise are qualitatively similar under different assumptions, our aim in this section is not to pin down the particular bias that would be generated by an employer's hiring decisions, but rather simply to demonstrate that such a bias might exist.

We first simulate which workers employers would hire if they only observed the characteristics on workers' resumes; we then simulate whom employers would hire if they additionally observed which workers had been referred. In each hiring scenario, we assume that employers want to maximize the fraction of workers who submit a response on a given day and that

[31] We do not have enough power to test the interaction of the relationship variables and treatment (in the individual experiment) or team type (in the team experiment). However, the coefficients on the interactions are not consistently signed across outcomes within each experiment.

they know the relationship between demographics and referral status, and performance.[32] 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 8 shows the results of the simple simulations. Results in the first row simulate hiring under the assumption that employers only see workers' resumes, not who was referred. To calculate a given worker's predicted performance, we first regress the performance of all other workers (excluding herself) on their resume characteristics and then use the estimated coefficients to predict the excluded worker's own performance. We also use this same predicted performance as a summary measure of workers' observable characteristics. Results in the second row simulate hiring under the assumption that employers observe not only workers' resume characteristics but also who was referred. We follow the same procedure to predict workers' performance except that the regressions of worker performance on observable characteristics also include an indicator for whether the worker was referred.

Panel A shows the fraction of referred and non-referred applicants that would have been hired under each scenario. If employers only took workers' resume characteristics into account, a higher fraction of referred (58 percent) than non-referred (39 percent) workers would have been hired because referred workers had better observable characteristics. However, if employers also observed who was referred, the fraction of referred applicants that would have been hired jumps to 77 percent; meantime, only 12 percent of non-referred applicants would have been hired. Panel B displays the summary measure of the hired workers' observable characteristics: when employers observe who was referred, hired non-referred workers have substantially better observable characteristics than hired referred workers.

Panel C shows the average actual submission rates of the referred and non-referred workers that would have been hired in each scenario. If employers did not observe who was referred, hired referred workers would have been substantially (19 percentage points) more likely to actually submit work. However, this difference would have been only five percentage points (and statistically indistinguishable from zero) if employers also observed who was referred.[33] This suggests that if we had only observed the performance of hired workers and did not observe all the characteristics employers used in hiring decisions, we might have mistakenly concluded that referrals contained little to no information about worker

[32] In practice, an employer may prefer to hire a referred worker over a non-referred 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.

[33] In fact, if we assume employers hired the top third or top quarter of the applicant pool (rather than the top half), hired referred workers would have performed two or three percentage points *worse* than hired non-referred workers when the employer used referral status in hiring decisions. As in the main specification, these differences are not significant.

performance.

8 Conclusion

This paper presents the results of three field experiments in an online labor market comparing the performance and turnover of referred and non-referred workers. Throughout, we find that even conditional on their resume characteristics, referred workers performed better and had less turnover than their non-referred counterparts. That is, referrals contained information about worker quality that was not present on workers' resumes.

Much of the performance and turnover differential between referred and non-referred workers was driven by selection. In the individual experiment, even non-monitored referred workers outperformed and outlasted non-referred workers. In the supplemental experiment, referred workers outperformed and outlasted non-referred workers even at a job for which they were not referred at a firm at which their referrers did not work.

However, we also find strong evidence that on-the-job interactions between referred workers and their referrers drove some of the performance differential. In particular, our results suggest that team production is an important benefit of referrals. Referred workers in the team experiment performed particularly well when working with their own referrers; they were also more eager to continue working in that pairing.

We find that workers referred by high-performers and workers with strong ties to their referrers performed particularly well. High-performers tended to refer workers who looked better on observables, but who performed better than expected even conditional on these characteristics. Referrals with strong ties to their referrers actually looked *worse* on paper than did those with weak ties. Nonetheless, it was the referrals with strong ties who performed better, even without conditioning on observable characteristics.

The existing literature finds mixed results on whether referred workers perform better than non-referred workers (e.g., Blau, 1990; Burks et al. 2013; Castilla 2005). We see our results as consistent with these seemingly divergent papers. The performance data in much of this literature come from workers firms chose to hire. However, if employers incorporate referrals into hiring decisions (for example, because referrals positively predict performance and persistence), hired referred workers could perform better than, worse than, or similarly to non-referred workers even though a referral is a positive signal of productivity.

We find that selection is important in explaining why referred workers outperform and outlast non-referred workers, but we have limited evidence on why. One explanation (as in Montgomery, 1991) is that there is simply homophily among friends: productive workers have productive friends. Another explanation is that (as in Beaman and Magruder, 2012) workers

have information on which of their friends are particularly productive and may choose these particularly productive workers to refer. Understanding the relative contributions of these two factors is an important question for future research.

References

Bandiera, O., I. Barankay, and I. Rasul (2005): "Social preferences and the response to incentives: Evidence from personnel data," *The Quarterly Journal of Economics*, 120(3), 917–962.

——— (2012): "Team incentives: Evidence from a firm level experiment," *CEPR Discussion Paper No. DP8776*.

Beaman, L., and J. Magruder (2012): "Who gets the job referral? Evidence from a social networks experiment," *The American Economic Review*, 102(7), 3574–3593.

Bentolila, S., C. Michelacci, and J. Suarez (2010): "Social contacts and occupational choice," *Economica*, 77(305), 20–45.

Bewley, T. F. (1999): *Why wages don't fall during a recession*. Harvard University Press.

Blau, G. (1990): "Exploring the mediating mechanisms affecting the relationship of recruitment source to employee performance," *Journal of Vocational Behavior*, 37(3), 303–320.

Brown, M., E. Setren, and G. Topa (2012): "Do informal referrals lead to better matches? Evidence from a firm's employee referral system," *Evidence from a Firm's Employee Referral System (August 1). FRB of New York Staff Report*, (568).

Bryan, G. T., D. Karlan, and J. Zinman (2012): "You can pick your friends, but you need to watch them: Loan screening and enforcement in a referrals field experiment," *NBER Working Paper 17883*.

Burks, S., B. Cowgill, M. Hoffman, and M. Housman (2013): "The facts about referrals: toward an understanding of employee referral networks," *Available at SSRN 2253738*.

Calvo-Armengol, A., and M. O. Jackson (2004): "The effects of social networks on employment and inequality," *The American Economic Review*, 94(3), 426–454.

Castilla, E. J. (2005): "Social networks and employee performance in a call center," *American Journal of Sociology*, 110(5), 1243–1283.

Costa, D. L., and M. E. Kahn (2003): "Cowards and heroes: Group loyalty in the American Civil War," *The Quarterly Journal of Economics*, 118(2), 519–548.

——— (2010): "Health, wartime stress, and unit cohesion: Evidence from Union Army veterans," *Demography*, 47(1), 45–66.

Datcher, L. (1983): "The impact of informal networks on quit behavior," *The Review of Economics and Statistics*, 65(3), 491–495.

Dustmann, C., A. Glitz, and U. Schönberg (2011): "Referral-based job search networks," *IZA Discussion Paper 5777*.

Fernandez, R. M., E. J. Castilla, and P. Moore (2000): "Social capital at work: Networks and employment at a phone center," *American Journal of Sociology*, 105(5), 1288–1356.

Fernandez, R. M., and N. Weinberg (1997): "Sifting and sorting: Personal contacts and hiring in a retail bank," *American Sociological Review*, pp. 883–902.

Granovetter, M. (1995): *Getting a job: A study of contacts and careers*. University of Chicago Press, Chicago.

Heath, R. (2013): "Why do firms hire using referrals?: evidence from Bangladeshi garment factories," *Working Paper*.

Herrmann, M. A., and J. E. Rockoff (2010): "Worker absence and productivity: Evidence from teaching," *NBER Working Paper 16524*.

Holzer, H. J. (1987): "Hiring procedures in the firm: Their economic determinants and outcomes," *NBER Working Paper 2185*.

Horton, J. J. (2013): "The effects of subsidizing employer search," *Working Paper*.

Ioannides, Y. M., and L. D. Loury (2004): "Job information networks, neighborhood effects, and inequality," *Journal of Economic Literature*, 42(4), 1056–1093.

Kugler, A. D. (2003): "Employee referrals and efficiency wages," *Labour Economics*, 10(5), 531–556.

Montgomery, J. D. (1991): "Social networks and labor-market outcomes: Toward an economic analysis," *The American Economic Review*, 81(5), 1408–1418.

oDesk Corporation (2013): *Slide deck with statistics on the oDesk marketplace*.

Pallais, A. (2014): "Inefficient hiring in entry-level labor markets," *American Economic Review*, forthcoming.

Petersen, T., I. Saporta, and M.-D. L. Seidel (2000): "Offering a job: meritocracy and social networks," *American Journal of Sociology*, 106(3), 763–816.

Pistaferri, L. (1999): "Informal networks in the Italian labor market," *Giornale degli Economisti e Annali di Economia*, pp. 355–375.

Rees, A. (1966): "Information networks in labor markets," *The American Economic Review*, 56(1/2), 559–566.

Simon, C. J., and J. T. Warner (1992): "Matchmaker, matchmaker: The effect of old boy networks on job match quality, earnings, and tenure," *Journal of Labor Economics*, 10(3), 306–330.

Topa, G. (2011): "Labor markets and referrals," *Handbook of Social Economics*, pp. 1193–1221.

Figure 1. oDesk Profile Example

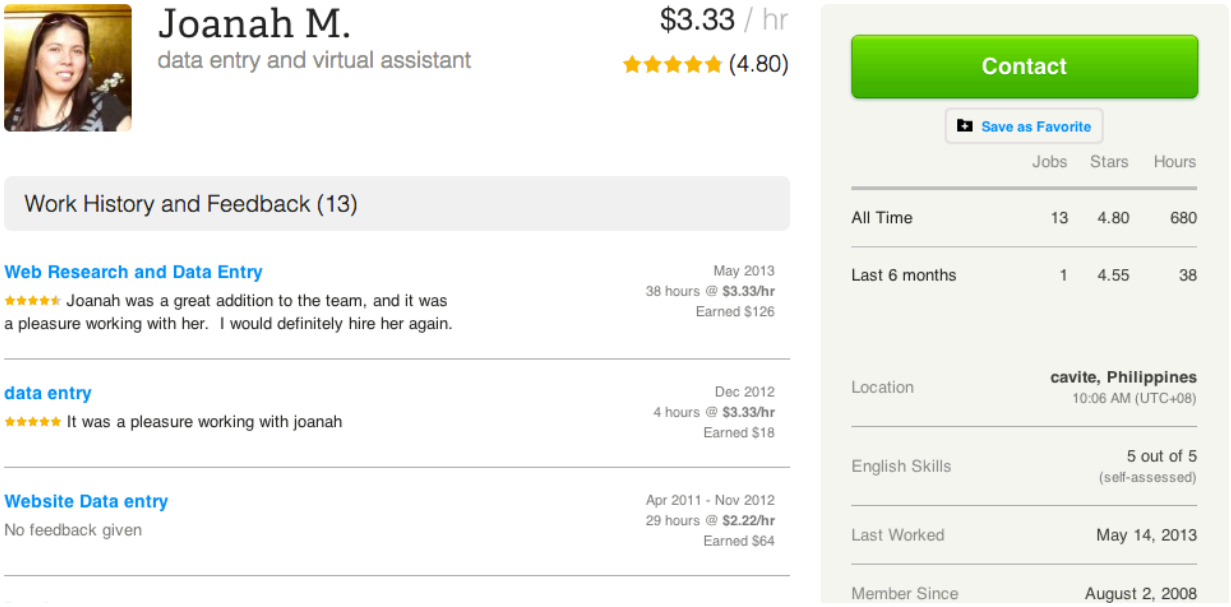


Figure 2. Submission Rates by Day, Individual Experiment

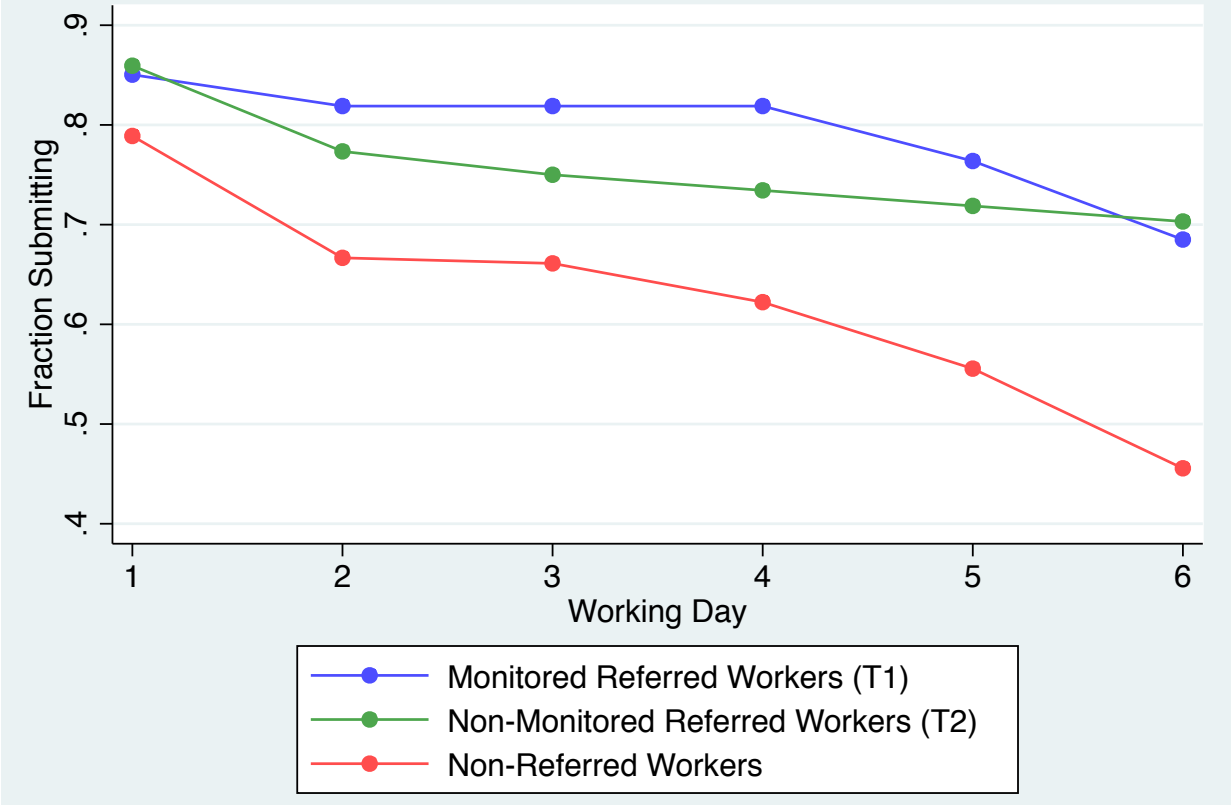


Table 1. Descriptive Statistics
Individual and Team Experiments

	All Referred Workers	Included Referred Workers	Included Non-Referred Workers	Difference
Has Prior Experience	0.62	0.64	0.28	**
Earnings	$1,481	$864	$353	**
Number of Previous Jobs	7.29	5.79	2.05	**
Has Feedback Score	0.53	0.55	0.24	**
Feedback Score	4.54	4.55	4.26	**
Posted Wage	$2.97	$2.58	$3.29	**
Days Since Joining oDesk	501	462	572	**
Has Portfolio	0.46	0.48	0.23	**
Number of Tests Passed	4.14	4.49	3.19	**
Has English Score	0.98	0.99	0.96	**
English Score	4.67	4.68	4.58	**
Agency Affiliated	0.18	0.12	0.06	**
Number of Degrees	1.27	1.35	1.01	**
Proposed Wage		$2.34	$2.59	**
Observations	1,854	537	274	

Notes: Each statistic in the table presents the mean of the characteristic indicated by the row for the sample indicated by the column. *All Referred Workers* denotes all workers who were referred, while *Included Referred Workers* is the subset of *All Referred Workers* who applied for our job and whom we hired for the individual or team experiment. *Included Non-Referred Workers* are non-referred workers who applied for our job and whom we hired for the individual or team experiment. *English Score* is self-reported English ability on a one-to-five scale, a *portfolio* is where a worker posts prior work, and *agency-affiliated* workers pay a fraction of their earnings to report they are part of a given group of oDesk workers (an agency). ** denotes that the means of the characteristic for *Included Referred Workers* and *Included Non-Referred Workers* are significantly different at the 5% level.

Table 2. Performance and Persistence
Individual Experiment: Base Group is Non-Monitored Referred Workers (Treatment 2)

A. All Days

	Submission	On-Time Submission	Accuracy	Re-Application
Monitored Referred (Treatment 1)	0.021	0.039	0.015	-0.033
	(0.042)	(0.047)	(0.040)	(0.035)
Non-Referred	-0.127**	-0.090*	-0.102**	-0.216**
	(0.046)	(0.048)	(0.042)	(0.043)
Base Group Mean (Treatment 2)	0.757	0.563	0.640	0.953
Controls	Yes	Yes	Yes	Yes
Observations	2,610	2,610	2,610	435
R-squared	0.078	0.063	0.075	0.130

B. Last Day Only, Controlling for Performance on First Five

	Submission	On-Time Submission	Accuracy	Re-Application
Monitored Referred (Treatment 1)	-0.055	0.000	-0.051	-0.047
	(0.044)	(0.053)	(0.039)	(0.035)
Non-Referred	-0.163**	-0.125**	-0.120**	-0.180**
	(0.048)	(0.050)	(0.043)	(0.041)
Daily Performance Controls	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	435	435	435	435
R-squared	0.528	0.405	0.535	0.306

Notes: Each column in each panel presents the results of a separate regression of the dependent variable (indicated by the column) on an indicator for being a referred worker in Treatment 1 and an indicator for being non-referred. In the first three columns in each panel, observations are worker-days and standard errors are clustered at the worker level. Regressions in Panel A include all six days of work while Regressions in Panel B are limited to observations on workers' last day of work. In the final column, observations are workers and Huber-White standard errors are presented. All regressions include the controls for worker characteristics listed in footnote 17. Regressions in Panel B add daily performance controls: each of the first three columns includes controls for the worker's performance as measured by the dependent variable on each of the first five days. The final column includes controls for each of the three performance measures on each of the six days. *, ** denote significance at the 10% and 5% levels, respectively.

Table 3. Performance and Persistence in New Firm
Supplemental Experiment: Base Group is All Referred Workers

A. All Workers

	Accepted Job Offer	Submission	On-Time Submission	Accuracy	Re-Application
Non-Referred	-0.064	-0.090*	-0.088*	-0.028	-0.121**
	(0.055)	(0.048)	(0.047)	(0.026)	(0.056)
Base Group Mean	0.678	0.518	0.499	0.247	0.553
Controls	Yes	Yes	Yes	Yes	Yes
Observations	435	2,175	2,175	2,175	435
R-squared	0.126	0.130	0.132	0.100	0.132

B. Conditional on Accepting Job Offer

	Submission	On-Time Submission	Accuracy	Re-Application
Non-Referred	-0.104*	-0.104*	-0.026	-0.131*
	(0.057)	(0.058)	(0.033)	(0.071)
Base Group Mean	0.763	0.735	0.363	0.815
Controls	Yes	Yes	Yes	Yes
Observations	1,325	1,325	1,325	265
R-squared	0.096	0.098	0.063	0.088

Notes: 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 non-referred worker. All regressions include the controls for worker characteristics listed in footnote 17. Panel A includes all workers to whom we made employment offers; Panel B includes only workers who accepted these offers. Observations in the first and last columns (*Accepted* , *Re-Application*) are workers; observations the middle three columns of regressions (*Submission* , *On-Time Submission* , *Accuracy*) are worker-days. *Accepted Job Offer* is an indicator for whether the worker accepted our invitation to work for the new firm created for the supplemental experiment. Huber-White standard errors are presented when observations are workers and standard errors are clustered at the worker level when observations are worker-days. *, ** denote significance at the 10% and 5% levels, respectively.

Table 4. Individual Diligence and Team Performance
Team Experiment: Base Group is Referred Workers Paired with Someone Else's Referrer (Type B)

A. Individual Diligence

	Logged in	Submitted	Individual Question Correct	Own Criteria in Slogan
Referred Worker When Working with Own Referrer (Type A)	0.018	0.046**	0.053*	0.004
	(0.017)	(0.017)	(0.029)	(0.034)
Non-Referred Worker When Working with Referrer (Type C)	-0.194**	-0.229**	-0.245**	-0.087
	(0.055)	(0.053)	(0.052)	(0.058)
Base Group Mean (Type B)	0.883	0.837	0.755	0.440
Controls	Yes	Yes	Yes	Yes
Observations	846	846	846	846
R-Squared	0.188	0.187	0.147	0.066

B. Team Performance

	Both Submitted	Team Question Matches	Same Slogan	Same Slogan & Both Criteria
Referred Worker and Own Referrer Team (Type A)	0.099**	0.287**	0.372**	0.103**
	(0.023)	(0.028)	(0.032)	(0.024)
Non-Referred Worker and Referrer Team (Type C)	-0.206**	-0.165**	-0.108**	-0.031
	(0.053)	(0.048)	(0.044)	(0.031)
Base Group Mean (Type B)	0.730	0.496	0.337	0.142
Controls	Yes	Yes	Yes	Yes
Observations	846	846	846	846
R-Squared	0.155	0.193	0.213	0.055

Notes: 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 non-referred workers (not referrers) are included. Observations in Panel B are at a team-PSA level. All regressions include the controls for worker characteristics listed in footnote 17. Standard errors are clustered at the blocking group level. *, ** denote significance at the 10% and 5% levels, respectively.

Table 5. Time Spent & Wanting to Partner Again, by Team Type
Team Experiment: Base Group is Referred Workers Paired with Someone Else's Referrer (Type B)

	A. Time Spent (Minutes)		B. Wants to Partner Again	
	Referrers	Referred & Non-Referred Workers	Referrers	Referred & Non-Referred Workers
Referred Worker Paired with Own Referrer (Type A)	5.922**	5.142**	0.556**	0.451**
	(1.752)	(1.559)	(0.030)	(0.033)
Non-Referred Worker Paired with Referrer (Type C)	1.135	-15.532**	-0.100**	0.009
	(1.445)	(3.121)	(0.041)	(0.060)
Constant	37.482	38.723	0.406	0.477
	(1.291)	(1.492)	(0.029)	(0.031)
Controls	No	No	No	No
Observations	846	846	717	612
R-squared	0.009	0.087	0.338	0.211

Notes: Each column in each panel reports the results of a separate regression of the dependent variable indicated by the panel title on indicators for being in a Type A team and for being in a Type C team. No controls are included. Observations are at a worker-PSA level. The first regression in each panel includes only referrers while the second includes only referred and non-referred workers. Standard errors are clustered at the blocking group level. ** denotes significance at the 5% level.

Table 6. Relationship between Referred Worker's Performance and Referrer's Performance
Individual and Supplemental Experiments

	Dependent Variable: Referred Worker's Submission Rate, Individual Experiment	Dependent Variable: Referred Worker's Submission Rate, Supplemental Experiment		
Referrer's Submission Rate, Individual Experiment	0.421**	0.246**	0.132	
	(0.066)	(0.079)	(0.082)	
Referred Worker's Submission Rate, Individual Experiment				0.409**
				(0.078)
Constant	0.456	0.331	0.222	0.201
	(0.059)	(0.065)	(0.421)	(0.063)
Controls	No	No	Yes	No
Observations	255	255	255	255
R-squared	0.192	0.034	0.184	0.087

Notes: Each column presents the results of a regression of the dependent variable indicated by the column on the independent variable indicated by the row. Each observation is a referred worker. Huber-White standard errors are in parenthesis. No controls are included except in the second-to-last column, which includes controls for referred worker characteristics listed in footnote 17. ** denotes significance at the 5% level.

Table 7. Relationship Strength, Observable Characteristics, and Performance
Individual, Supplemental, and Team Experiments: Referred Workers

A. Observable Characteristics

	Tests Passed	Number of Assignments	Feedback Score
Relationship Strength Index	-0.462**	-1.456**	-0.038
	(0.138)	(0.579)	(0.030)
Constant	4.447	5.774	4.545
	(0.125)	(0.593)	(0.041)
Controls	No	No	No
Observations	532	532	293
R-squared	0.025	0.011	0.003

B. Performance, No Controls

	Submission (Individual Experiment)	Submission (Supplemental Experiment)	Same Slogan (Team Experiment)
Relationship Strength Index	0.041**	0.018	0.053**
	(0.020)	(0.027)	(0.020)
Constant	0.780	0.518	0.514
	(0.021)	(0.029)	(0.022)
Controls	No	No	No
Observations	1,512	1,260	560
R-squared	0.012	0.003	0.009

C. Performance, With Controls

	Submission (Individual Experiment)	Submission (Supplemental Experiment)	Same Slogan (Team Experiment)
Relationship Strength Index	0.047**	0.037	0.051**
	(0.020)	(0.026)	(0.022)
Controls	Yes	Yes	Yes
Observations	1,512	1,260	560
R-squared	0.106	0.155	0.048

Notes: Each column in each panel reports the results of a separate regression of the dependent variable indicated by the column on an index for the strength of the referrer-referred worker relationship. This index is defined in Section 6 of the text and has mean zero and standard deviation one. All regressions in the table include only referred workers. Regressions in Panel A include referred workers in both the individual and team experiments. Observations are at the worker level. No controls are included; Huber-White standard errors are in parentheses. Regressions in Panel B include no controls, while regressions in Panel C include the controls for worker characteristics listed in footnote 17. The first two columns of Panels B and C include workers from only the individual and supplemental experiments, respectively. In these columns, outcomes are observed at the worker-day level and standard errors are clustered by worker. The final column of Panels B and C includes only workers from the team experiment; outcomes are observed at the worker-PSA level and standard errors are clustered by blocking group. ** denotes significance at the 5% level.

Table 8. Simulated Hiring
Individual Experiment: Assuming Top 50% of Applicants Hired

	A. Fraction Hired		B. Measure of Observables		C. Actual Submission Rate		
	Referred Applicants	Non-Referred Applicants	Hired Referred Workers	Hired Non-Referred Workers	Hired Referred Workers	Hired Non-Referred Workers	Difference
Observe Characteristics Only	58%	39%	80%	79%	82%	64%	19%**
Observe Characteristics & Referral Status	77%	12%	77%	86%	81%	75%	5%
Applicant Pool Average			73%	68%	81%	75%	15%**

Notes: The first row simulates hiring under the assumption that employers observed only workers' resume characteristics, but not their referral status. The second row simulates hiring assuming employers observed workers' resume characteristics and referral status. Panel A presents the fraction of referred and non-referred workers that would have been hired under each scenario. Panel B presents a summary measure of the observables of the referred and non-referred workers who would have been hired. (This is the average predicted probability of submission, based on the observable characteristics listed in footnote 17. It is described in the text.) Panel C presents the actual submission rate of the referred and non-referred workers who would have been hired. The column labeled *Difference* provides the difference in average submission rates of the referred and non-referred workers who would have been hired under each scenario. ** denotes that this difference is significant at the 5% level.

9 Appendix A: Communication in the Team Experiment

Here we discuss communication differences across team types. In light of our findings, for robustness we also re-estimate performance differences by team type with the inclusion of several additional controls.

Panel A of Appendix Table 9 shows how the team types differed in their communication methods. We regress each communication outcome on indicators for being in Type A and Type C teams; as before, the base group is Type B teams. Controls for the characteristics of referred and non-referred workers are included throughout. The first column considers chat box use, defined as both partners typing at least one message in the chat box. The second column considers the total number of messages sent by both partners during the task and is limited to teams that used the chat box. Because we directly observe what is written in the chat box, both of these measures are known for all teams and do not rely on worker reports.

The last two columns consider communication outside the chat box, such as on Skype. When workers submitted their slogans for each task, we asked if they had used other forms of communication. We code teams as using other forms of communication if at least one partner reported doing so. The third column addresses selection into answering this question. Here we regress a dummy for whether at least one teammate answered this question on team type. In the final column, we regress an indicator for reporting using other forms of communication on team type. This final specification includes only teams that answered the communication question.

Type A teams communicated the most, both in and out of the chat box. Relative to Type B teams, Type A teams were slightly, though insignificantly, more likely to use the chat box. When they did use the chat box, Type A teams sent about one-third more messages. The biggest difference between the communication of Type A and Type B teams, however, is in the frequency with which they used other forms of communication. While 38 percent of Type B teams reported using other forms of communication, Type A teams were twice as likely to do so; the magnitude of this difference implies that the difference itself cannot be driven by the small difference in the likelihood of answering this question.[34]

On the other hand, Type C teams were significantly less likely to use the chat box. This is not surprising since the chat box was on the site and non-referred workers were significantly less likely ever to log in.

We have observed that, relative to Type B teams, Type A teams communicated more

[34] At least one partner answered this question in 95 percent of Type B teams; Type B teams were slightly more likely to answer this question than either of the other team types.

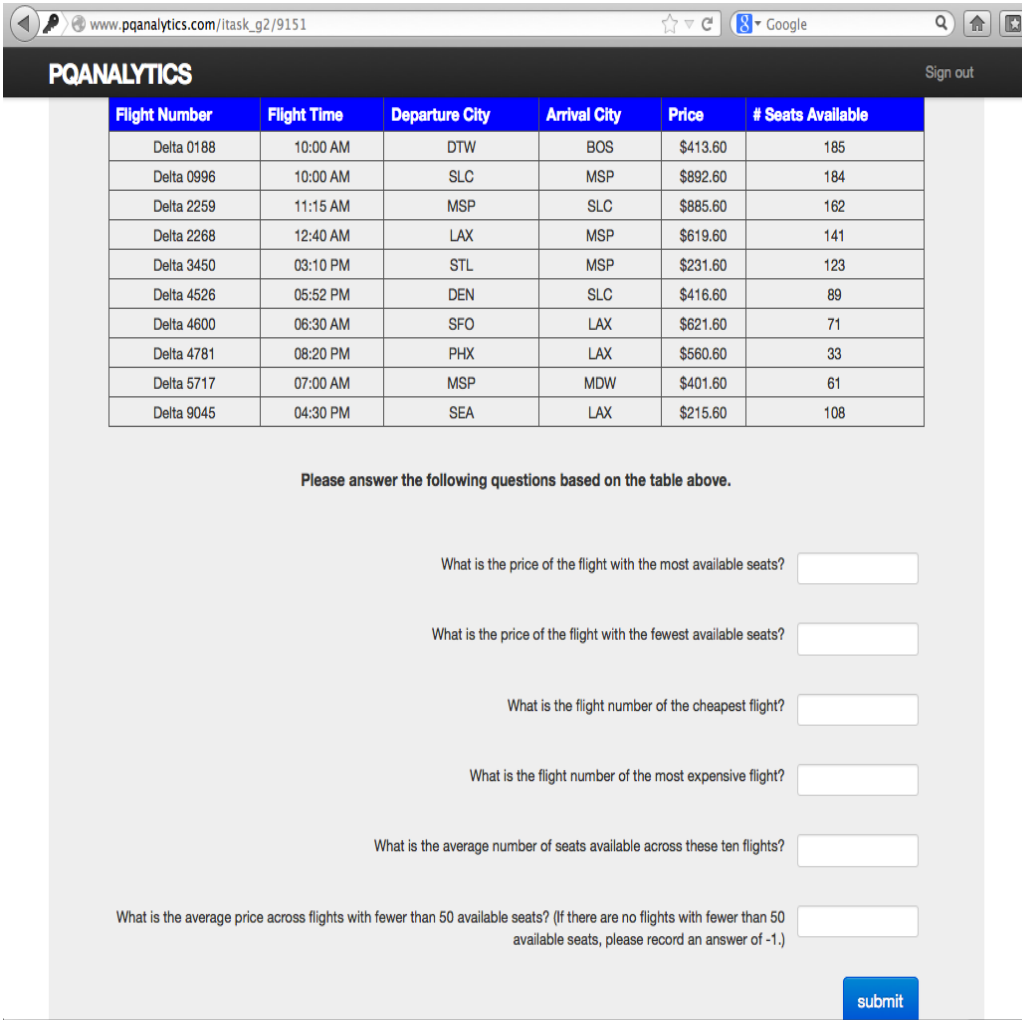
both in and out of the chat box. They also spent more time on the task (Table 5). Panel B of Appendix Table 9 provides evidence that, even controlling for communication and time spent, Type A teams still outperformed Type B teams. We replicate the main team performance specifications with controls for referred and non-referred workers' characteristics (from Panel B of Table 4), adding as controls an indicator for using the chat box, the number of messages sent in the chat box, an indicator for using other methods of communication, and the number of minutes spent by each partner.[35]

Unsurprisingly, more communication and more time spent both led to better outcomes. For example, teams that sent the median number of messages in the chat box (eight) were 21 percentage points more likely to answer the team question the same way and 15 percentage points more likely to provide the same slogan than were teams that did not use the chat box, all else equal. Teams in which each partner spent an additional five minutes each on the task were, all else equal, three percentage points more likely to have their team question match and two percentage points more likely to submit the same slogan.

Even conditioning on the type of communication used, number of messages sent, and minutes spent by each partner, however, Type A teams remained 14 percentage points more likely to provide the same answer to the team question and 23 percentage points more likely to submit the same slogan than Type B teams. Type C teams, meanwhile, remained substantially less likely to do either (11 percentage points and seven percentage points, respectively).

[35] If neither partner answered the question about using other forms of communication, we set the indicator for having reported communication outside the chat box to zero. Thus, this dummy also directly captures the effect of having at least one partner submit work.

Appendix Figure 1. Individual Experiment Task Site, Referred and Non-Referred Workers



Flight Number	Flight Time	Departure City	Arrival City	Price	# Seats Available
Delta 0188	10:00 AM	DTW	BOS	$413.60	185
Delta 0996	10:00 AM	SLC	MSP	$892.60	184
Delta 2259	11:15 AM	MSP	SLC	$885.60	162
Delta 2268	12:40 AM	LAX	MSP	$619.60	141
Delta 3450	03:10 PM	STL	MSP	$231.60	123
Delta 4526	05:52 PM	DEN	SLC	$416.60	89
Delta 4600	06:30 AM	SFO	LAX	$621.60	71
Delta 4781	08:20 PM	PHX	LAX	$560.60	33
Delta 5717	07:00 AM	MSP	MDW	$401.60	61
Delta 9045	04:30 PM	SEA	LAX	$215.60	108

Appendix Figure 2. Performance Report Example

Me
Feb 01

Dear [redacted],

We are writing to update you, as promised, on the performance of a colleague you referred.

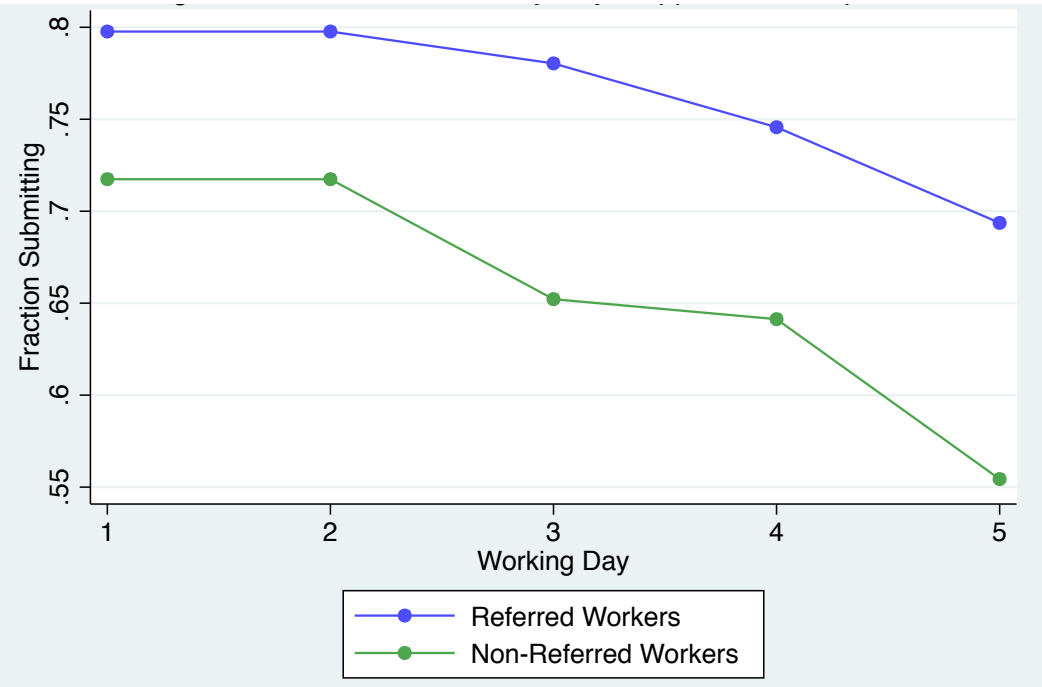
Yesterday,
[redacted] did not submit their work.
[redacted] did not submit their work by 11AM PHT.
[redacted] did not answer all of the questions.

Cumulative Performance :
* Percentage of days work was submitted : 80%
* Percentage of days work was submitted by 11AM PHT : 60%
* Percentage of days all questions were answered : 80%

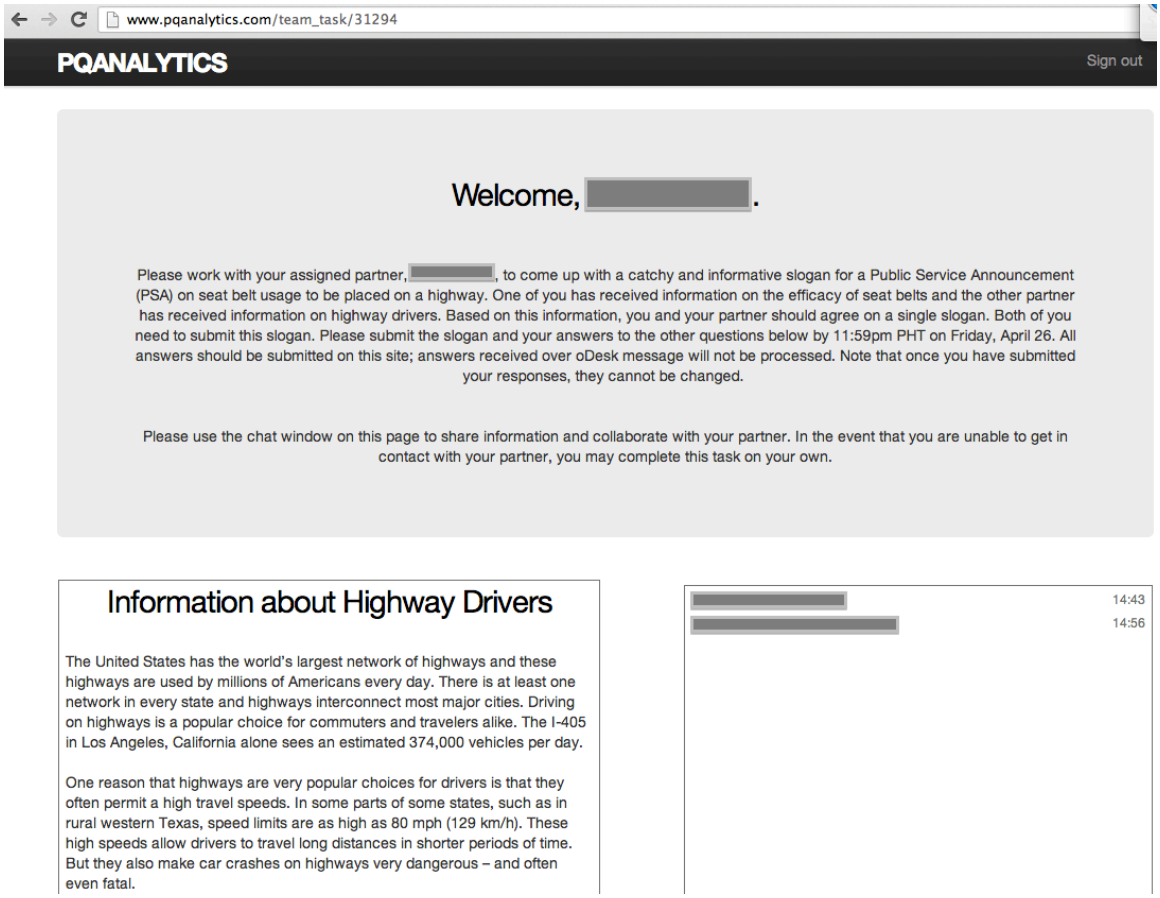
Sincerely,
The Hiring Team

Appendix Figure 3. Supplemental Experiment Task Site

Appendix Figure 4. Submission Rates by Day, Supplemental Experiment



Appendix Figure 5. Team Experiment Task Site



Appendix Table 1. Descriptive Statistics

Individual and Team Experiments, Workers Asked to Refer

	Referred Someone	Referred No One	Difference	Included Referrers
Has Prior Experience	1.00	1.00		1.00
Earnings	$2,917	$2,397	**	$2,932
Number of Previous Jobs	12.58	11.07	**	12.35
Has Feedback Score	1.00	1.00		1.00
Feedback Score	4.80	4.80		4.81
Posted Wage	$2.84	$2.77	*	$2.85
Days Since Joining oDesk	689	709		666
Has Portfolio	0.69	0.61	**	0.69
Number of Tests Passed	5.80	5.60		5.84
Has English Score	1.00	0.99	**	1.00
English Score	4.79	4.79		4.77
Agency Affiliated	0.25	0.24		0.21
Number of Degrees	1.40	1.35		1.41
Proposed Wage	$2.50	$2.51		$2.51
Observations	1,246	1,867		455

Notes: Each statistic in the table presents the mean of the characteristic indicated by the row for the sample indicated by the column. *Referred Someone* denotes workers who referred at least one other worker to our firm, whether or not we hired that worker. *Referred No One* denotes workers who referred no workers to our firm. *Included Referrers* is a subset of *Referred Someone* and includes only those workers whose referral we hired. *English Score* is self-reported English ability on a one-to-five scale, a *portfolio* is where a worker posts prior work, and *agency-affiliated* workers pay a fraction of their earnings to report they are part of a given group of oDesk workers (an agency). *, ** denotes that the means of the characteristic for *Referred Someone* and *Referred No One* are significantly different at the 10% and 5% levels, respectively.

Appendix Table 2. Randomization Assessment
Individual Experiment

	Referrers		Referred Workers	
	Treatment 1	Treatment 2	Treatment 1	Treatment 2
Has Prior Experience	1.00	1.00	0.73	0.75
Earnings	$2,919	$2,996	$1,396	$1,379
Number of Previous Jobs	12.78	13.09	8.28	10.14
Has Feedback Score	1.00	1.00	0.62	0.64
Feedback Score	4.80	4.76	4.66	4.59
Posted Wage	$2.78	$2.85	$2.68	$2.72
Days Since Joining oDesk	645	676	489	566
Has Portfolio	0.64	0.68	0.47	0.50
Number of Tests Passed	5.78	5.78	4.98	5.31
Has English Score	1.00	1.00	0.98	1.00
English Score	4.84	4.79	4.75	4.66
Agency Affiliated	0.08	0.08	0.05*	0.10*
Number of Degrees	1.50	1.36	1.34	1.51
Proposed Wage	$2.53	$2.53	$2.40	$2.37
Observations	86	87	127	128

Notes: Each cell presents the mean of the characteristic indicated by the row for the sample indicated by the column. Only workers in the individual experiment are included. *English Score* is self-reported English ability on a one-to-five scale, a *portfolio* is where a worker posts prior work, and *agency-affiliated* workers pay a fraction of their earnings to report they are part of a given group of oDesk workers (an agency). * denotes the Treatment 1 and Treatment 2 group means are statistically different at the 10% level.

Appendix Table 3. Performance and Persistence, With Different Controls
Individual Experiment: Base Group is Non-Monitored Referred Workers (Treatment 2)

A. All Days, No Controls

	Submission	On-Time Submission	Accuracy	Re-Application
Monitored Referred	0.036	0.053	0.034	-0.032
(Treatment 1)	(0.042)	(0.047)	(0.039)	(0.030)
Non-Referred	-0.132**	-0.079*	-0.101**	-0.225**
	(0.042)	(0.045)	(0.039)	(0.038)
Constant	0.757	0.563	0.640	0.953
	(0.031)	(0.034)	(0.028)	(0.019)
Controls	No	No	No	No
Observations	2,610	2,610	2,610	435
R-squared	0.027	0.013	0.020	0.085

B. All Days, with Second Order Controls

	Submission	On-Time Submission	Accuracy	Re-Application
Monitored Referred	0.012	0.043	0.015	-0.027
(Treatment 1)	(0.042)	(0.046)	(0.040)	(0.039)
Non-Referred	-0.143**	-0.080	-0.109**	-0.184**
	(0.050)	(0.052)	(0.046)	(0.054)
Controls	Yes	Yes	Yes	Yes
Second Order Controls	Yes	Yes	Yes	Yes
Observations	2,610	2,610	2,610	435
R-squared	0.190	0.160	0.180	0.270

C. Last Day Only, Second Order Controls

	Submission	On-Time Submission	Accuracy	Re-Application
Monitored Referred	-0.045	0.023	-0.032	-0.041
(Treatment 1)	(0.049)	(0.057)	(0.043)	(0.039)
Non-Referred	-0.161**	-0.104*	-0.090*	-0.125**
	(0.054)	(0.057)	(0.046)	(0.050)
Daily Performance Controls	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Second Order Controls	Yes	Yes	Yes	Yes
Observations	435	435	435	435
R-squared	0.617	0.510	0.625	0.452

Notes: Panel A replicates Panel A of Table 2, eliminating the controls for worker characteristics. Panels B and C replicate Panels A and B, respectively, of Table 2 with additional control variables. In addition to the characteristics listed in footnote 17, these regressions all include *Second Order Controls:* the square of each non-binary characteristic in footnote 17 and the interaction of each pair of characteristics in footnote 17. *, ** denote significance at the 10% and 5% levels, respectively.

Appendix Table 4. Selection into Accepting Job Offer in the Supplemental Experiment
Data from the Individual and Supplemental Experiments

	Dependent Variable: Accepted Job Offer in Supplemental Experiment			
	Submission	On-Time Submission	Accuracy	Re-Application
Performance in Individual Experiment × Non-Referred	0.131	0.133	0.219	0.008
	(0.129)	(0.123)	(0.135)	(0.143)
Performance in Individual Experiment	0.164*	0.113	0.159	0.217*
	(0.096)	(0.081)	(0.099)	(0.119)
Non-Referred	-0.122	-0.114	-0.165*	-0.026
	(0.106)	(0.086)	(0.098)	(0.136)
Controls	Yes	Yes	Yes	Yes
Observations	435	435	435	435
R-squared	0.154	0.144	0.161	0.149

Notes: Each column presents the results of a separate regression of an indicator for a worker accepting our job offer in the supplemental experiment on a measure of her performance in the individual experiment, an indicator for being a non-referred worker, and the interaction of these two indicators. Each column uses a different performance measure indicated by the column heading. Observations are workers; only referred and non-referred workers are included. Each regression contains controls for the individual characteristics listed in footnote 17. * denotes significance at the 10% level.

Appendix Table 5. Performance and Persistence in New Firm, with Different Controls
Supplemental Experiment: Base Group is All Referred Workers

A. All Workers, No Controls

	Accepted Job Offer	Submission	On-Time Submission	Accuracy	Re-Application
Non-Referred	-0.167**	-0.182**	-0.178**	-0.079**	-0.236**
	(0.047)	(0.042)	(0.042)	(0.022)	(0.047)
Constant	0.678	0.518	0.499	0.247	0.553
	(0.029)	(0.029)	(0.029)	(0.015)	(0.031)
Controls	No	No	No	No	No
Observations	435	2,175	2,175	2,175	435
R-squared	0.029	0.033	0.031	0.019	0.055

B. All Workers, with Second Order Controls

	Accepted Job Offer	Submission	On-Time Submission	Accuracy	Re-Application
Non-Referred	-0.042	-0.073	-0.073	-0.020	-0.097
	(0.064)	(0.055)	(0.055)	(0.029)	(0.064)
Controls	Yes	Yes	Yes	Yes	Yes
Second Order Controls	Yes	Yes	Yes	Yes	Yes
Observations	435	2,175	2,175	2,175	435
R-squared	0.268	0.242	0.239	0.201	0.292

C. Conditional on Accepting Job Offer, with Second Order Controls

	Submission	On-Time Submission	Accuracy	Re-Application
Non-Referred	-0.119*	-0.118*	-0.047	-0.198**
	(0.067)	(0.070)	(0.038)	(0.090)
Controls	Yes	Yes	Yes	Yes
Second Order Controls	Yes	Yes	Yes	Yes
Mean of Dependent Variable				
Base Group (Referred Workers)	0.763	0.735	0.363	0.815
Observations	1,325	1,325	1,325	265
R-squared	0.247	0.25	0.192	0.358

Notes: Panel A replicates Panel A of Table 3, eliminating the controls for worker characteristics. Panels B and C replicate Panels A and B, respectively, of Table 3 with additional control variables. In addition to the characteristics listed in footnote 17, these regressions all include *Second Order Controls* : the square of each non-binary characteristic in footnote 17 and the interaction of each pair of characteristics in footnote 17. *, ** denote significance at the 10% and 5% levels, respectively.

Appendix Table 6. Performance in Team Experiment, with Different Controls
Team Experiment: Base Group is Referred Workers Paired with Someone Else's Referrer (Type B)

A. Individual Diligence , No Controls

	Logged in	Submitted	Individual Question Correct	Own Criteria in Slogan
Referred Worker When Working with Own Referrer (Type A)	0.018	0.046**	0.053*	0.004
	(0.017)	(0.017)	(0.028)	(0.033)
Non-Referred Worker When Working with Referrer (Type C)	-0.294**	-0.312**	-0.287**	-0.138**
	(0.051)	(0.048)	(0.049)	(0.052)
Controls	No	No	No	NO
Observations	846	846	846	846
R-Squared	0.124	0.134	0.102	0.018

B. Individual Diligence, with Second Order Controls

	Logged in	Submitted	Individual Question Correct	Own Criteria in Slogan
Referred Worker When Working with Own Referrer (Type A)	0.018	0.046**	0.053*	0.004
	(0.018)	(0.018)	(0.030)	(0.035)
Non-Referred Worker When Working with Referrer (Type C)	-0.131**	-0.176**	-0.193**	-0.044
	(0.058)	(0.056)	(0.057)	(0.051)
Controls	Yes	Yes	Yes	Yes
Second Order Controls	Yes	Yes	Yes	Yes
Observations	846	846	846	846
R-Squared	0.420	0.388	0.309	0.213

C. Team Performance, No Controls

	Both Submitted	Team Question Matches	Same Slogan	Same Slogan & Both Criteria
Referred Worker and Own Referrer Team (Type A)	0.099**	0.287**	0.372**	0.103**
	(0.023)	(0.028)	(0.032)	(0.024)
Non-Referred Worker and Referrer Team (Type C)	-0.280**	-0.206**	-0.142**	-0.053*
	(0.048)	(0.047)	(0.041)	(0.029)
Controls	No	No	No	No
Observations	846	846	846	846
R-Squared	0.117	0.164	0.194	0.031

D. Team Performance, with Second Order Controls

	Both Submitted	Team Question Matches	Same Slogan	Same Slogan & Both Criteria
Referred Worker and Own Referrer Team (Type A)	0.099**	0.287**	0.372**	0.103**
	(0.024)	(0.030)	(0.034)	(0.025)
Non-Referred Worker and Referrer Team (Type C)	-0.162**	-0.075	-0.032	0.026
	(0.061)	(0.054)	(0.054)	(0.031)
Controls	Yes	Yes	Yes	Yes
Second Order Controls	Yes	Yes	Yes	Yes
Observations	846	846	846	846
R-Squared	0.314	0.327	0.311	0.170

Notes: Panels A and B replicate Panel A of Table 4 eliminating the controls for worker characteristics (Panel A) and adding *Second Order Controls* (Panel B). Panels C and D replicate Panel B of Table 4, eliminating the controls for worker characteristics (Panel C) and adding *Second Order Controls* (Panel D). The *Second Order Controls are* the square of each non-binary characteristic in footnote 17 and the interaction of each pair of characteristics in footnote 17. *, ** denote significance at the 10% and 5% levels, respectively.

Appendix Table 7. Relationship Between Referrer Characteristics and Referred Worker Characteristics
Individual and Team Experiments

	Dependent Variable: Referred Worker Characteristic										
	Earnings	Number of Previous Jobs	Feedback Score	Posted Wage	Days Since Joining oDesk	Has Portfolio	Number of Tests Passed	English Score	Agency Affiliated	Number of Degrees	Proposed Wage
Referrer Characteristic	0.060*	0.095*	0.204	0.231**	0.264**	0.272**	0.157**	0.285**	0.317**	0.075*	0.367**
	(0.034)	(0.052)	(0.180)	(0.049)	(0.049)	(0.044)	(0.043)	(0.068)	(0.050)	(0.044)	(0.052)
Constant	688.6	4.637	3.562	1.928	289.95	0.296	3.580	3.319	0.060	1.238	1.425
	(113.3)	(0.685)	(0.867)	(0.146)	(33.337)	(0.035)	(0.277)	(0.333)	(0.011)	(0.069)	(0.134)
Observations	537	537	296	537	537	537	537	533	537	537	537
R-Squared	0.014	0.009	0.005	0.036	0.060	0.064	0.032	0.057	0.146	0.006	0.103

Notes: Each column reports the results of regressing the value of an observable characteristic for a referred worker on the value of the same characteristic for her referrer. Each column corresponds to a different characteristic, indicated by the column header. All 537 hired referred workers from the individual and team experiments are included, although *Feedback Score* and *English Score* are missing for some workers. *English Score* is self-reported English ability on a one-to-five scale, a *portfolio* is where a worker posts prior work, and *agency-affiliated* workers pay a fraction of their earnings to report they are part of a given group of oDesk workers (an agency). Huber-White standard errors are in parenthesis. *, ** denote significance at the 10% and 5% levels, respectively.

Appendix Table 8. Characteristics of the Referrer-Referred Worker Relationship
Individual and Team Experiments

	Included Referred Workers	Excluded Referred Workers
How Well Referrer Knows Referral		
1 (Hardly at all)	1%	2%
2	2%	2%
3	5%	3%
4	14%	9%
5	20%	19%
6 (Extremely Well)	57%	65%
Observations	535	1,314
Frequency of Interaction		
Less than Once a Month	2%	4%
About Once a Month	5%	4%
Less than Weekly, More than Monthly	8%	4%
About Once a Week	13%	9%
Less than Daily, More than Weekly	21%	14%
About Once a Day	19%	16%
More than Once a Day	32%	47%
Observations	533	1,311
Number of People Known in Common		
0 to 4	21%	18%
5 to 9	16%	16%
10 to 19	16%	18%
20 to 29	11%	10%
30 or more	37%	39%
Observations	535	1,314
Sometimes Work in Same Room	0%	44%
Observations	537	1,317

Notes: This table presents the distributions of referrers' responses to questions about their relationships with their referrals for two different samples, indicated by the column headings. *Included Referred Workers* are referred workers we hired in either the individual or team experiment. *Excluded Referred Workers* are workers who were referred to us, but whom were not included in either experiment.

Appendix Table 9. Team Communication and Performance Controlling for Communication and Time Spent
Team Experiment: Base Group is Referred Workers Paired with Someone Else's Referrer (Type B)

A. Communication by Team Type

	Chat Box Use	Total Chat Messages (Conditional on Use)	Answered Communication Question	Reported Outside Communication
Referred Worker and Own	0.025	4.346*	-0.028*	0.376**
Referrer Team (Type A)	(0.042)	(2.544)	(0.016)	(0.040)
Non-Referred Worker and	-0.090*	0.805	-0.032	-0.043
Referrer Team (Type C)	(0.047)	(2.706)	(0.027)	(0.038)
Controls	Yes	Yes	Yes	Yes
Mean of Dependent Variable				
Base Group (Type B)	0.408	13.522	0.947	0.378
Observations	846	307	846	778
R-Squared	0.047	0.062	0.017	0.193

B. Team Performance Controlling for Communication and Time Spent

	Both Submitted	Team Question Matches	Same Slogan	Same Slogan & Both Criteria
Referred Worker and Own	-0.028	0.140**	0.225**	0.025
Referrer Team (Type A)	(0.026)	(0.031)	(0.037)	(0.029)
Non-Referred Worker and	-0.144**	-0.112**	-0.066*	-0.014
Referrer Team (Type C)	(0.035)	(0.041)	(0.038)	(0.029)
Used Chat Box	0.223**	0.167**	0.116**	0.020
	(0.028)	(0.037)	(0.041)	(0.041)
Total Chat Messages	0.001*	0.005**	0.006**	0.003*
	(0.001)	(0.001)	(0.001)	(0.002)
Used Outside Communication	0.236**	0.320**	0.336**	0.173**
	(0.030)	(0.033)	(0.042)	(0.031)
Minutes Spent by Referrer	0.003**	0.002**	0.002**	0.002**
	(0.001)	(0.001)	(0.001)	(0.000)
Minutes Spent by Referred or	0.005**	0.003**	0.002**	0.001**
Non-Referred Worker	(0.001)	(0.001)	(0.000)	(0.000)
Controls	Yes	Yes	Yes	Yes
Mean of Dependent Variable				
Base Group (Type B)	0.730	0.500	0.337	0.142
Observations	846	846	846	846
R-Squared	0.497	0.478	0.467	0.171

Notes: 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 are at the worker-PSA level. *Chat Box Use* is an indicator for whether each partner typed at least one message in the chat box. *Total Chat Messages* is the aggregate number of messages sent between the two partners, and is conditional on chat box use. *Answered Communication Question* is an indicator for whether at least one partner responded to the question at the end of that task about how the partners had communicated. *Reported Outside Communication* is an indicator for whether either partner reported communicating using methods other than the chat box and is conditional on at least one partner having answered the communication question. Regressions in both panels control for the characteristics of referred and non-referred workers listed in footnote 17. Regressions in Panel B also control for whether the team used the chat box, the number of chat messages sent, whether either partner reported using other forms of communication, and (separately) the number of minutes spent by both partners. Standard errors are clustered at the blocking group level. *, ** denote significance at the 10% and 5% levels, respectively.