The Effects of Algorithmic Labor Market Recommendations: Evidence from a Field Experiment

John J. Horton, *New York University*

Algorithmically recommending workers to employers for the purpose of recruiting can substantially increase hiring: in an experiment conducted in an online labor market, employers with technical job vacancies that received recruiting recommendations had a 20% higher fill rate compared to the control. There is no evidence that the treatment crowded out hiring of nonrecommended candidates. The experimentally induced recruits were highly positively selected and were statistically indistinguishable from the kinds of workers employers recruit "on their own." Recommendations were most effective for job openings that were likely to receive a smaller applicant pool.

I. Introduction

The rise of the Internet created a hope among economists and policy makers that it would lower labor market search costs and lead to better market outcomes. Evidence for whether this hope was fulfilled is mixed (Kuhn

I am grateful to the oDesk Research and oDesk team members for encouragement, ideas, and assistance in this project. Larry Katz, Richard Zeckhauser, Luke Stein, Jacob Leshno, Mandy Pallais, Tal Gross, Peter Coles, Ramesh Johari, and especially Dana Chandler offered numerous helpful comments and suggestions. Thanks to seminar participants at Harvard Business School, Stanford Management Science and Engineering, Wharton, INSEAD, London Business School, and New York University Stern. Thanks to Robin Yerkes Horton, Carolyn Yerkes, Ada Yerkes Horton, Heidi Yerkes, and David Yerkes for help in preparing the manuscript. Views expressed in this paper are my own. Contact the author at john.horton@stern.nyu.edu. Information concerning access to the data used in this article is available as supplementary material online.

[*Journal of Labor Economics*, 2017, vol. 35, no. 2]

Submitted January 30, 2014; Accepted March 16, 2016; Electronically published January 26, 2017
DOI: 10.1086/689213

and Skuterud 2004; Kuhn and Mansour 2014), but to date the rise of the Internet seems to have had modest effects on search and matching in the conventional labor market. However, simply providing parties with searchable listings of jobs and resumes—the core functionality of online job boards—hardly exhausts the possibilities created by marketplace digitization.

In many online product markets, the creating platform now goes beyond simply providing information but rather makes explicit, algorithmically generated recommendations about whom to trade with or what to buy (Resnick and Varian 1997; Adomavicius and Tuzhilin 2005; Varian 2010). Algorithmic recommender systems can try to infer preferences, determine the feasible choice set, and then solve the would-be buyer's constrained optimization problem. At their best, algorithmic recommendations can incorporate information not available to any individual party. Furthermore, these recommendations have zero marginal cost, and recommendation quality potentially improves with scale.

To date, algorithmic recommendations have been rare in labor markets, but as more aspects of the labor market become computer-mediated, recommendations will become increasingly feasible. However, it is not clear that labor market recommendations can meaningfully improve upon what employers can do for themselves. Perhaps choosing who is appropriate for a particular job opening requires evaluating ineffable qualities that are difficult to capture in a statistical model. Or perhaps assembling a pool of reasonable applicants is simply not that costly to employers. Beyond the perspective of the individual employer, a concern with recommendations is that, by design, they encourage an employer to consider some workers but not others. If crowd-out effects are strong—which has been the case in some job search assistance programs in conventional labor markets (Crépon et al. 2013)—recommendation interventions are less attractive from a social welfare perspective.

In this paper, I report the results of an experimental intervention in which algorithmically generated recommendations were made to employers about which workers to recruit for their job openings.[1] The context for the experiment was oDesk, a large online labor market. On oDesk, employer recruiting is one of two "channels" employers use to get applicants for their job openings—the other is to rely on "organic" applicants finding the job listing and applying without prompting by the employer. Before the experiment, employers could recruit only by searching through listings of workers and inviting those who looked promising.

I find that when offered algorithmic recommendations, a large fraction of employers follow them: the treatment increased the fraction of employers

[1] I use the term "employer" throughout the paper for consistency with the extant labor literature rather than as a commentary on the precise nature of the contractual relationship between parties on the platform.

recruiting by nearly 40%. Recruited workers accepted invitations and applied for the associated job at the same rates in both the treatment and control groups. As such, the treatment substantially increased the number of recruited applicants in the applicant pools of treated employers.

Recruited applicants were highly positively selected in terms of market experience and past earnings. This characterization held in both the treatment and control groups—experimentally induced recruits "looked like" the kinds of workers employers recruit on their own. Employers showed a strong preference for screening recruited applicants relative to nonrecruited organic applicants, but this difference did not depend on the treatment assignment. This lack of a difference across experimental groups undercuts the notion that employers believed that recommended workers were better or worse than their observable characteristics suggested.

Being offered recruiting assistance raised the probability that an employer hired someone for his/her job opening, but this was not the case for all types of job openings: the treatment increased the overall fill rate in technical job openings by 20%, but it had no detectable effect on nontechnical job openings. The strong effects on technical job openings do "show up" in the entire sample, in that the treatment substantially raised the probability that the wage bill for a job opening exceeds $500 (technical job openings, when filled, lead to projects that are, on average, larger than nontechnical projects in terms of wage bill).

There are several potential reasons why the treatment was only effective for technical job openings, but the most likely explanation is that (i) technical job openings attract fewer organic applicants, which are substitutes for recruited applicants and (ii) employers with technical openings seem to value experience and are less cost-sensitive than their nontechnical counterparts—and recruited applicants tend to be both more experienced and more expensive. Highlighting the importance of the organic applicant count in explaining treatment effect heterogeneity, when the treatment is conditioned on the expected number of organic applicants to a job opening, I find that the technical/nontechnical distinction is largely explained by differences in applicant pool size: the treatment is more effective for jobs expected to receive few applicants than for jobs expected to receive many applicants.

Despite raising the fill rate for technical job openings, there is no evidence of crowd-out of organic applicants for those job openings. The likely explanation is that the low hire rate—less than 50% of job openings are filled—creates "space" to increase hiring without crowd-out. Those matches that were formed in the treatment group were indistinguishable from matches in the control group with respect to match outcomes, such as the total wage bill and feedback score. However, the study is underpowered to detect even fairly large changes in match quality.

To summarize, the evidence is most parsimoniously explained by the following: (i) employers acted upon the recommendations because it was cheap to do so and the recommended candidates were similar to the kind of workers they would have recruited themselves—namely, relatively high-cost but high-quality applicants; (ii) recommendations were effective at raising fill rates for technical job openings because these employers have relatively high returns to additional high-quality applicants; (iii) where they were effective, recommendations had little crowd-out because the baseline vacancy fill rate was low enough that market expansion effects could dominate.

The main implication of the paper is that a relatively unsophisticated algorithmic recommender system—unsophisticated compared to what oDesk (now known as "Upwork") does presently—can substitute for some of the work employers have to do when filling a job opening, and this substitution can substantially increase hiring. However, the paper also highlights the economic nature of the employer's hiring problem; recommendation efficacy turned out to depend less on the details of the recommendation algorithm—at least in terms of the kinds of recruits it generated—and more on how employers valued recruited applicants and how many organic applicants they could expect to receive in the absence of recommendations. Although we can imagine algorithms that might improve match quality—some standardized job tests, which are a kind of algorithm for hiring seem successful at doing so—the algorithm used in this paper worked primarily by expanding the pool of potential applicants by lowering the employer's costs of assembling such a pool.

This intervention was conducted in a setting where search costs are presumably quite low: oDesk is information rich in that both sides have access to the universe of job seekers and job openings and at every step both sides have comprehensive data on past job histories, wages, feedback scores, and so on. In this environment, one might suppose that the marginal benefit of algorithmic recommendations would be low, and yet this is strongly not the case. In conventional settings where the stakes are higher, one might expect employers to expend more effort in recruiting and screening, but this implies that the opportunity for reducing costs is even greater, even if the expected benefits in terms of match formation might be lower.

II. Empirical Context

During the past 15 years, a number of online labor markets have emerged. In these markets, firms hire workers to perform tasks that can be done remotely, such as computer programming, graphic design, data entry, research, and writing. Markets differ in their scope and focus, but common services provided by the platforms include maintaining job listings, hosting user profile pages, arbitrating disputes, certifying worker skills, and maintaining feedback systems. The experiment described in this paper was conducted on oDesk, the largest of these online labor markets.

In the first quarter of 2012, $78 million were spent on oDesk. The 2011 wage bill was $225 million, representing 90% year-on-year growth from 2010. As of October 2012, more than 495,000 employers and 2.5 million freelancers had created profiles, though a considerably smaller fraction were active on the site. Approximately 790,000 job openings were posted in the first half of 2012. See Agrawal et al. (2015) for additional descriptive statistics on oDesk.

Based on dollars spent, the top skills in the marketplace are technical skills, such as web programming, mobile applications development (e.g., iPhone and Android), and web design. Based on hours worked, the top skills are web programming again, but also data entry, search engine optimization, and web research, which are nontechnical and require little advanced training. The difference in the top skills based on dollars versus hours reflects a fundamental split in the marketplace between technical and nontechnical work. There are highly skilled, highly paid freelancers working in nontechnical jobs, yet a stylized fact of the marketplace is that technical work tends to pay better, generate longer-lasting relationships, and require greater skill.

There has been some research that focuses on the oDesk marketplace. Pallais (2014) shows via a field experiment that past worker experience on oDesk is an excellent predictor of being hired for subsequent work on the platform. Stanton and Thomas (2012) use oDesk data to show that agencies (which act as quasi-firms) help workers find jobs and break into the marketplace. Agrawal, Lacetera, and Lyons (2013) investigate what factors matter to employers in making selections from an applicant pool and present some evidence of statistical discrimination; that paper also supports the view of employers selecting from a more-or-less complete pool of applicants rather than serially screening.

A. Job Posting, Recruiting, Screening, and Hiring on oDesk

The process for filling a job opening on oDesk is qualitatively similar to the process in conventional labor markets. First, a would-be employer on oDesk creates a job post: he/she writes a job title and describes the nature of the work, chooses a contractual form (hourly or fixed-price), and specifies what skills the project requires (both by listing skills and choosing a category from a mutually exclusive list) and what kinds of applicants he/she is looking for in terms of past experience. Employers also estimate how long the project is likely to last. Once the job post is written, it is reviewed by oDesk and then posted to the marketplace.

Once posted to the marketplace, would-be job applicants can view all the employer-provided job post information. Additionally, oDesk also presents verified attributes of the employer, such as their number of past jobs, average paid wage rate, and so on. When a worker applies to a job opening, he/she offers a bid (which is an hourly wage or a fixed price, depending on contract type) and includes a cover letter. After applying, the applicant immediately appears in the employer's ATS, or "applicant tracking system." The employer can view an applicant's first name, profile picture, offered bid, and a few pieces of other oDesk-verified information, such as hours worked and his/her feedback rating from previous projects (if any). Employers can click on a worker's application to view his/her full profile, which has that worker's disaggregated work history, with per-project details on feedback received, hours worked, and earnings. As these clicks are recorded by oDesk, they provide an intermediate measure of employer interest in a particular applicant.

Although all job applications start with the worker applying to a job opening, not all of these applications are initiated by the worker: as in conventional labor markets, employers on oDesk may choose to recruit candidates to apply for their jobs. Employer recruiting on oDesk begins with the employer searching for some skill or attribute he/she is looking for in candidates; the search tools on oDesk will return lists of workers and will contain information about that worker's past work history. The employer can "invite" any worker he/she is interested in recruiting. These recruiting invitations are not job offers, but rather invitations to apply to the employer's already-posted job opening. As will become evident, these recruited applicants tend to be highly positively selected: they have more experience, higher past wages, greater earnings, and so forth, and consequently, they also bid considerably more for hourly jobs than nonrecruited organic applicants.

Of course recruited workers are not required to apply to the job opening—only about half do apply. Those who do apply appear in the employer's ATS alongside whatever organic applicants the job opening has attracted. Employers are free to evaluate candidates at any time after they post their jobs. Presumably different employers use different approaches to evaluate applicants depending upon their urgency in filling the job opening and the fixed costs of a screening "session." Anecdotally, some employers screen applicants as they arrive, while others wait to process them in batch, after a suitable number have arrived.

Although employers can recruit at any time, employers generally recruit shortly after posting their job openings, before they receive any organic applicants. If recruiting is costly, a natural question is why would employers ever recruit "ex ante," that is, before receiving organic applicants? One reason is that it allows the employer to obtain a better applicant pool more quickly, and given that most employers want to fill their job openings as soon as possible, ex ante recruiting can be rational. Ex ante recruiting also allows employers to evaluate candidates "in batch" by assembling a more or less complete pool of applicants first and then screening them all at once.

If the employer makes a hire, oDesk intermediates the relationship. If the project is hourly, hours worked are measured via custom tracking software that workers install on their computers. The tracking software, or "Work

Diary," essentially serves as a digital punch clock, allowing hours worked and earnings to be measured essentially without error.

B. Competing Markets

The oDesk marketplace is not the only marketplace for online work (or IT work more generally). As such, one might worry that every job opening on oDesk is simultaneously posted on several other online labor market sites and in the conventional market. If this were the case, it would make interpreting events happening on oDesk more complex, particularly for an experiment focused on raising the number of matches formed; perhaps any observed increase in the fill rate simply came at the expense of some other marketplace that is unobserved.

Despite the possibility of simultaneous posting, survey evidence suggests that online and offline hiring are only very weak substitutes and that "multi-homing" of job openings on other online labor markets is relatively rare. When asked what they would have done with their most recent project if oDesk were not available, only 15% of employers responded that they would have made a local hire. In this same survey, online employers report that they are generally deciding among (*a*) getting the work done online, (*b*) doing the work themselves, and (*c*) not having the work done at all. The survey also found that 83% of employers said that they listed their last job job openings on oDesk alone. This self-report appears to be credible, as Horton (2015) found limited evidence of multi-homing when comparing jobs posted on oDesk and its largest (former) rival, Elance. This limited degree of multi-homing narrows the scope of potential crowd-out effects from marketplace interventions to those happening within the platform rather than across platforms.

III. Description of the Experiment

In June 2011, oDesk launched an experimental feature that targeted new employers, with "new" defined as those who had not previously posted a job opening on oDesk. Immediately after posting a job opening, a treated employer was shown up to six recommended workers that the employer could recruit to apply for his/her job opening. Control employers received the status quo experience of no recommendations. The total sample for the experiment consisted of 6,209 job openings, which is the universe of job openings that were posted by new employers during the experimental period and for which recommendations could be made (regardless of treatment assignment). The randomization was effective, and the experimental groups were well balanced (see the appendix for details).[2]

[2] This appendix also discusses the possibility of cross-unit effects, i.e., violations of the SUTVA assumption and discusses why, given the size of the experiment relative to the market as a whole, such concerns are likely unwarranted.

The actual recommendations were delivered to treated employers via a pop-up interface, a screen-shot of which is shown in figure 1. From this interface, the employer could compare each recommended worker's photograph, listed skills, average feedback score, and stated hourly wage. If the employer clicked on a worker's application in the ATS, he/she could see that worker's country, total hours worked on the platform, passed skills tests, past employer evaluations, and other pieces of potentially match-relevant information. Employers could choose to invite any number of the recommended workers to apply for their jobs (including no one at all). Once a treated employer closed the recommendations pop-up window, he/she experienced the same interface and opportunities as employers in the control group.

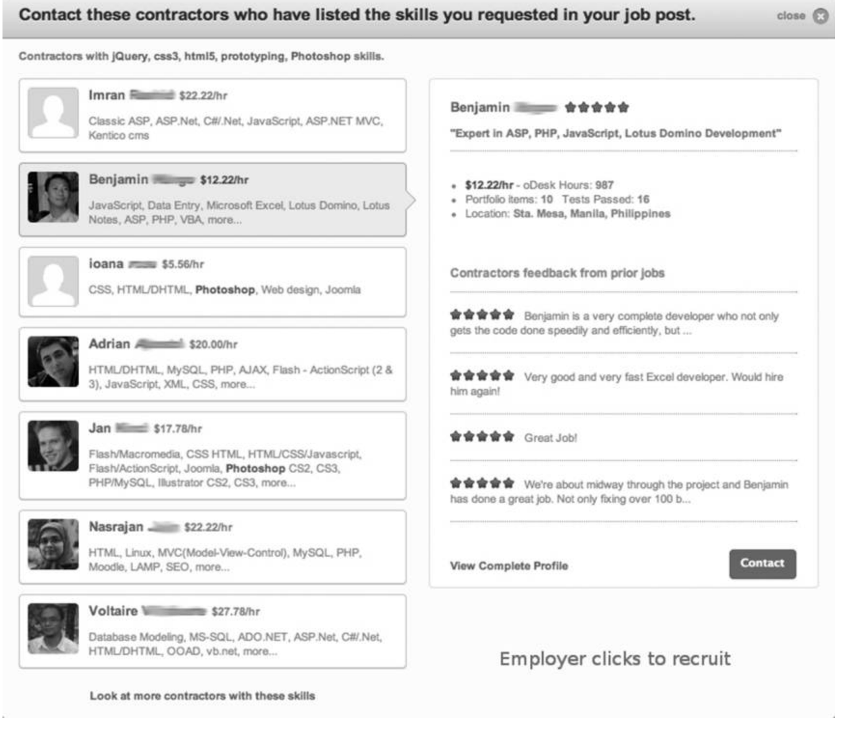

FIG. 1.—Recommendations shown to treated employers after posting their job openings. This figure shows the interface presented to employers in the treatment group. It displays a number of recommended workers with good on-platform reputations, skills relevant to the employer's job opening, and predicted availability for the employer's project.

Recommendations were made based on the fit of a statistical model using historical oDesk hiring data. The model incorporated measures of worker relevance to the job in question, the worker's ability, and the worker's availability, that is, capacity to take on more work. Relevance was measured by the degree of overlap in the skills required for the job opening and the skills listed by the worker in his/her profile. Ability was defined as a weighted sum of the applicant's skill test scores, feedback ratings, and past earnings. Availability was inferred from signals, such as the worker recently ending a project or applying to other job openings. If the employer invited a recommended worker, the experience of the invited worker and employer were from then on identical to what would have happened had the employer simply found and invited that worker on his/her own. The invited worker did not know that the invitation was experimentally induced, nor was the employer later notified of that worker's invited status in the ATS.

One unfortunate limitation of the design of the experiment was that the identity of workers who were recommended (or would have been recommended in the control) was not logged. For some outcomes of interest, such as the overall job fill rate, not having access to the individual recommendations is irrelevant. However, for other questions, such as whether recommendations were "better" in some categories of work, not having the actual recommendations is a limitation. As a work-around, it is possible to make a reasonable inference about which invitations and follow-on applications were more likely to be experimentally induced: because recommendations were presented as a pop-up immediately after a job opening was posted, recruiting invitations made shortly after the posting were more likely to be caused by the treatment. For analyses where it is useful to identify experimentally induced invitations and applications, I define "ex ante" recruiting as recruiting that occurred within the first hour after the associated job opening was posted.

IV. Conceptual Framework and Related Work

An employer's search and screening process could benefit from an algorithmic approach in several ways. For example, an algorithm could be used to help evaluate an already assembled pool of candidates. Using "algorithms" to screen candidates has a long (uncomputerized) history: standardized exams, like those used for entrance into a civil service, take applicant characteristics—namely, their answers to exam questions—and return a recommendation about whom to hire.[3] Information technology has made this kind of test-based screening easier, as it simplifies administration and grading—Autor and Scarborough (2008) describe a large national retail chain switching from an informal screening process to a standardized testing pro-

[3] The Chinese civil service examination system dates back to the mid-600s.

cedure, finding that the move increased productivity and increased tenure at the firm. Hoffman, Kahn, and Li (2015), looking at the staggered introduction of pre-employment testing, also find that testing improves match quality and tenure. Furthermore, those managers exercising discretion—overruling the algorithm—end up with systematically worse outcomes.

Another use of algorithms could simply be to identify and deliver "reasonable" applicants for a job opening, who the employer could then recruit at his/her discretion. The algorithm could do a better or worse job than a firm's unassisted search efforts could accomplish, with algorithm performance ranging from (*a*) a random sample of candidates in the market to (*b*) the precise applicants, out of all possible applicants, that would make the best matches if hired. Interestingly, even a random sample might be valuable, as a perennial concern in labor markets is that those actively looking for work are adversely selected (such as through the mechanism in Gibbons and Katz [1991]). A sample of potential recruiting targets might be especially welcome in labor markets where parties generally have not had ready access to the whole pool of potential workers.

The experiment described in this paper was intended to increase the applicant pool size, with the hope that the recommendations would be good enough that the employer would not simply discard them. As such, the experiment was most conceptually similar to active labor market policies in which a government agency assists with job finding, with the focus in these programs typically on helping workers. These programs tend to have positive, albeit modest, effects on employment probability (Card, Kluve, and Weber 2010; Kluve 2010). However, a perennial concern with such programs is that the benefits mainly come at the expense of those not assisted. This concern is highlighted by Gautier et al. (2012) and was recently illustrated by Crépon et al. (2013), which was a large-scale, job-finding assistance program that seemed to "work" mostly by displacing nonassisted job seekers.

The labor market intervention in this paper was demand-focused, with assistance offered to employers. As such, understanding how this assistance might help requires a model of employer search and screening. Unfortunately, there is relatively little research in economics on how employers fill job openings and thus little guidance on how worker-finding assistance should affect outcomes (Oyer and Schaefer 2011). Most of the literature on labor market search has focused on workers searching for jobs, not firms searching for workers. An exception is Barron and Bishop (1985), which finds that employers with hard-to-fill job openings or those that require more training report screening larger pools of applicants and screening each applicant more intensively. Pellizzari (2011) finds that more intensive recruitment by a sample of British employers is associated with better-quality matches. The resultant matches pay more, last longer, and lead to greater employer satisfaction, though the direction of causation is not clear.

A. Models of Employer Search

Existing models of employer search are similar to simple job search models (e.g., Barron and Bishop 1985; Barron, Black, and Loewenstein 1989; Burdett and Cunningham 1998); firms serially screen applicants who arrive over time without recall or firms hire the first applicant above some "reservation value" for the position. These models are hard to map to key empirical features of the oDesk domain, such as the decision whether or not to recruit and whether the treatment would have heterogeneous effects based on the nature of the job opening. Furthermore, as van Ours and Ridder (1992) points out, there is little empirical basis for this sequential search-based view of hiring; they find that "almost all vacancies are filled from a pool of applicants that is formed shortly after the posting of the vacancy."

Given the difficulty of mapping existing employer search models to the current domain, I develop a simple model more closely tied to the empirical context and that can help interpret the experimental results. Some of the modeling choices are motivated by the experimental results, so the empirical results should not be interpreted as ex post tests of the model. Rather, the model is intended to offer one way of framing what the experimental intervention did and to see if the various empirical results can at least be rationalized by a simple model.

I model the employer's decision about recruiting intensity as his/her weighing the cost of recruiting versus the benefit from recruiting in terms of a better applicant pool and thus a higher probability of filling his/her job. I consider how a change in the cost of recruiting affects the extent of employer recruiting and the probability that a job opening is filled; a focus is on whether differences in organic applicant counts lead to heterogeneous treatment effects. I also characterize how a reduction in recruiting costs affects the probability that a nonrecruited applicant is hired (which speaks directly to the crowd-out question).

B. The Employer's Recruiting and Hiring Problem

A firm is trying to produce some output, y, that will yield py in the product market when sold. The firm knows that it will get a collection of A organic applicants for sure. It can also recruit R applicants, at a cost of cR. Both kinds of workers are ex ante homogeneous,[4] but each worker varies in how good of a "fit" he/she is for a particular job and thus how likely he/she is to produce the output. The firm has to pay a hired worker the market wage of w.

[4] This homogeneity assumption is counterfactual in the actual data—recruited candidates tend to be highly positively selected—but this fact does not change the essential features of the firm's recruiting problem. However, accounting for it in a model would substantially increase model complexity.

Once a worker applies, the firm observes $\hat{y}$, which is the firm's unbiased estimate of the probability that the worker can produce the output. Assume that $\hat{y} \sim U[0,1]$. To keep things simple, I assume that the firm only has the time and capacity to screen one and only one applicant and that the pay-off to hiring is large enough that screening the top candidate is always worthwhile. The firm selects the applicant with the highest predicted productivity, $\hat{y}^*$, and puts the candidate through additional screening to see if he/she can actually produce the output. As the firm's estimates are unbiased, the probability that the firm makes a hire following this additional screening is

$$\Pr\{\text{Hire} \mid \hat{y}^*\} = \hat{y}^*, \tag{1}$$

with probability $1 - \hat{y}^*$ the firm hires no one. The firm's screening technology is perfect, and so a hired worker can produce the output with certainty.

The expected productivity of the top applicant, conditioned upon having an applicant pool of size $A + R$, is

$$\mathbf{E}[\hat{y}^* \mid A, R] = \int_0^1 (A + R)\hat{y}^{A+R-1} d\hat{y} = \frac{A + R}{A + R + 1}. \tag{2}$$

The firm's recruiting optimization problem is thus

$$\max_R \, (p - w)\mathbf{E}[\hat{y}^* \mid A, R] - cR, \tag{3}$$

and the first-order condition is $p - w = (A + R + 1)^2 c$, so the interior solution is

$$R^* = \sqrt{(p - w)/c} - 1 - A. \tag{4}$$

The optimal recruiting solution has appealing comparative statics: there is more recruiting when recruiting is cheaper and more recruiting when the profit earned from producing the output is greater. As one would expect given that there is nothing special about recruits, the number of organic applicants, A, enters linearly, with a unit coefficient in equation (4), implying that an increase of one additional organic applicant would cause the firm to want to decrease recruited applicants by exactly one. If A is sufficiently large, a corner solution would result, with $R^* = 0$.

If c is exogenously lowered, such as by a third party making recruiting recommendations, then R^* goes up, as

$$\frac{\partial R^*}{\partial c} = \frac{\partial}{\partial c}\left[\sqrt{(p - w)/c}\right] < 0. \tag{5}$$

Note that the effect on R^* from a small change in c does not depend on A, but rather only $p - w$ and c. The effect of more recruits on the probability a hire is made is

$$\frac{\partial \hat{y}^*}{\partial R} = \frac{1}{(A + R + 1)^2} > 0. \tag{6}$$

When R increases, $\hat{y}^*$ increases, and thus more hires are observed. However, this increase in the hire probability from more recruits is declining in the number of organic applicants, as

$$\frac{\partial}{\partial A}\left[\frac{\partial \hat{y}^*}{\partial R}\right] = -\frac{2}{(A + R + 1)^3} < 0. \tag{7}$$

For job openings with a lower A, the effect of a lower c will have a larger effect on hiring probabilities than in job openings with a higher A. The change in the probability of a hire is

$$\frac{\partial \hat{y}^*}{\partial c} = \frac{\partial \hat{y}^*}{\partial R^*}\frac{\partial R^*}{\partial c}. \tag{8}$$

In terms of crowd-out, the expected number of hired organic applicants per job opening is $A/(A + R)\hat{y}^*$, which is just the fraction of applicants who are organic times the fill probability, $\hat{y}^*$. The change in this expectation from a change in c, is

$$\frac{\partial(\Pr\{\text{hired organic applicant}\})}{\partial c} = -\frac{A}{(A + R)^2}\hat{y}^*\frac{\partial R^*}{\partial c}. \tag{9}$$

The gross amount of crowd-out of organic applicants caused by a reduction in c is the change in recruiting, scaled by the base hire rate.[5]

V. Results

In the language of the model presented above, an interpretation of the recommendations treatment was that it lowered c, the cost of recruiting. In the model, a lowered cost of recruiting should increase recruiting and raise the probability that a hire is made, though the size of the treatment effect on the hire rate depends on the number of organic applicants. Many of the experimental results are apparent in a simple comparison of outcome means across the treatment and control groups: table 1 reports the fraction of employers that recruited, made a hire, and ultimately spent more than \$500 against their job opening, by experimental group. The "made a hire" outcome is further decomposed into an indicator for whether the employer hired a recruited applicant or an organic applicant. The top panel of the table uses all job openings as the sample, while the middle and bottom panels show results for nontechnical and technical job openings, respectively.

[5] Note that the envelope theorem allows us to ignore the marginal jobs that are induced to fill by the treatment.

Table 1
Outcome Means by Experimental Groups

	Treatment Mean: $\bar{X}_{\mathrm{Trt}}$	Control Mean: $\bar{X}_{\mathrm{Ctl}}$	Difference in Means: $\bar{X}_{\mathrm{Trt}} - \bar{X}_{\mathrm{Ctl}}$	p-Value	Significance Level
Outcomes for all vacancies:					
Recruited	.207	.148	.059		***
	(.007)	(.006)	(.010)	< .001	
Made a hire	.295	.276	.019		
	(.008)	(.008)	(.011)	.102	
Hired an organic applicant	.272	.266	.006		
	(.008)	(.008)	(.011)	.592	
Hired recruited applicant	.048	.040	.009		+
	(.004)	(.003)	(.005)	.097	
Wage bill > $500	.072	.058	.014		*
	(.005)	(.004)	(.006)	.023	
Outcomes for nontechnical vacancies:					
Recruited	.199	.153	.045		***
	(.010)	(.009)	(.014)	.001	
Made a hire	.331	.339	−.008		
	(.012)	(.012)	(.017)	.630	
Hired an organic applicant	.311	.323	−.012		
	(.012)	(.012)	(.017)	.476	
Hired recruited applicant	.045	.047	−.003		
	(.005)	(.005)	(.008)	.730	
Wage bill > $500	.059	.049	.009		
	(.006)	(.005)	(.008)	.254	
Outcomes for technical vacancies:					
Recruited	.216	.143	.073		***
	(.011)	(.009)	(.014)	< .001	
Made a hire	.259	.216	.043		**
	(.011)	(.010)	(.015)	.005	
Hired an organic applicant	.233	.212	.022		
	(.011)	(.010)	(.015)	.146	
Hired recruited applicant	.052	.032	.020		**
	(.006)	(.004)	(.007)	.006	
Wage bill > $500	.085	.066	.019		*
	(.007)	(.006)	(.009)	.040	

NOTE.—This table reports outcome means and standard errors (in parentheses) across experimental groups from the recommendations experiment. In the experiment, the treatment group of employers received algorithmically generated recommendations of candidates for their vacancies, while the control group did not. The unit of randomization was the posted vacancy. The top panel means are calculated using all observations. The middle panel uses only nontechnical vacancies, while the bottom panel uses only technical vacancies. Reported p-values are the for two-sided t-tests of the null hypothesis of no difference in means across groups. "Recruited" is an indicator for whether the employer made an invitation to a worker within 1 hour of posting their job. "Filled vacancy" is an indicator for whether the employer hired a worker and spent more than $1. A "recruited applicant" is one who applied to the vacancy after being recruited by the employer, whereas an organic applicant is one who applied without prompting by the employer. The "wage bill" is the total amount spent by the employer against that vacancy.

+ $p \leq .10$.
* $p \leq .05$.
** $p \leq .01$.
*** $p \leq .001$.

The key experimental results are that (i) the treatment increased recruiting among all job openings; (ii) experimentally induced recruited applicants were similar on observables to the workers employers recruited on their own; (iii) the treatment increased fill rates among technical job openings only; (iv) where recommendations were effective, there was no observable reduction in hiring of nonrecommended organic applicants. These results are explored in more depth in the subsections that follow. The effects of the treatment on match quality are presented in Appendix Section A2.3 due to space considerations. The main conclusion from Appendix Section A2.3 is that within the limits of statistical power available, there is no difference across experimental groups on match outcomes such as feedback and the total wage bill.

A. Employer ex ante Recruiting

The treatment was designed to increase employer recruiting. Table 2 shows that this goal was achieved for all categories of work, with recruiting increasing substantially among treated employers. There is no evidence that the treatment increased ex post recruiting, that is, recruiting that occurred more than 1 hour after the job opening was posted, undercutting the notion that the treatment may have worked simply by alerting the employer to the

Table 2
Effects of the Recommendations Treatment on Employer ex ante Recruiting

	Measures of Employer Recruiting				
	Any ex ante?		Any ex post?	Number ex ante	
	(1)	(2)	(3)	(4)	(5)
Assigned to treatment (Trt)	.059***	.045***	−.004	.151***	.094
	(.010)	(.014)	(.008)	(.032)	(.135)
Technical vacancy (Tech)		−.010			
		(.013)			
Trt × Tech		.028			
		(.019)			
Constant	.148***	.153***	.124***	.332***	2.384***
	(.006)	(.009)	(.006)	(.020)	(.100)
Sample no. of recruits > 1?	No	No	No	No	Yes
Sample no. of recruits < 11?	No	No	No	Yes	Yes
Observations	6,209	6,209	6,209	6,127	1,018
R^2	.006	.006	.00003	.004	.0005

NOTE.—This table reports ordinary least squares regressions where the dependent variable is some measure of recruiting. The same sample is used in both cols. 1 and 2, and it consists of all job openings assigned to the recommendations experiment. Ex ante recruiting is the employer recruiting within the first hour after posting a job opening. "Treatment" is an indicator that the employer received recommendations about candidates to recruit. "Technical" is an indicator for that opening requiring computer programming. The col. 3 outcome is an indicator for any ex post recruiting (i.e., recruiting after the first hour of a job posting). In col. 4, the outcome is the count of ex ante recruiting invitations, but with the sample restricted to job openings sending 10 or fewer of these invitations. Column 5 has the same outcome, but with the sample further restricted to those employers sending at least one ex ante recruiting invitation. Robust standard errors are reported in parentheses.

*** $p \le .001$.

possibility of recruiting. The treatment appears to have worked primarily on the extensive margin (i.e., increasing the number of employers recruiting at all) rather than the intensive margin (i.e., the number of candidates recruited, conditional upon recruiting).

Column 1 of table 2 reports an estimate of

$$\text{EmpRecruited}_j = \beta_0 + \beta_1 \text{Trt}_j + \varepsilon, \tag{10}$$

where EmpRecruited_j is an indicator for whether the employer recruited ex ante and Trt_j is whether the employer received the recommendations treatment. The treatment increased ex ante recruiting by about 6 percentage points, from a base of only about 15 percentage points.

To test whether the treatment differed by the nature of the work, column 2 reports an estimate of

$$\text{EmpRecruited}_j = \beta_0 + \beta_1 \text{Trt}_j + \beta_2 \text{Tech}_j + \beta_3 \left(\text{Trt}_j \times \text{Tech}_j\right) + \varepsilon, \tag{11}$$

where Tech_j is an indicator for whether the job opening was technical. The coefficient on the interaction term is positive and potentially economically significant—uptake was about 3 percentage points higher for technical job openings—but the coefficient is not statistically significant.

If the treatment "worked" by simply alerting employers to the possibility of recruiting, then the treatment might have a positive effect on ex post recruiting. However, this does not appear to be the case. In column 3, the outcome is an indicator for whether the employer sent any ex post recruiting invitations, which I define as all invitations sent after the first hour of job posting. The coefficient is negative, small in magnitude, and precisely estimated.

In addition to measuring recruiting as a binary measure, I also examine the effect of the treatment on the count of recruited applicants. Column 4 reports a regression where the dependent variable is the number of ex ante recruits per job opening. The coefficient on the treatment is positive and highly significant, with the treatment increasing the count of recruits from a base of about a third of an ex ante recruit per opening to nearly half an ex ante recruit per opening. Note that for this estimation, the sample is restricted to job openings with 10 or fewer ex ante recruiting invitations in order to improve the precision of the estimates. (A small number of employers send very large numbers of recruiting invitations.)

In the column 5 regression, the sample is further restricted to job openings where the employer sent at least one ex ante recruiting invitation. This allows me to test whether the treatment increased the number of recruits, conditional upon recruiting at least one recruit. Although the coefficient on treatment indicator is positive, it is not significant, and it is small in magnitude relative to the baseline count in the control, which is a little less than 2.5 ex ante recruits per opening.

The main effect of the treatment was to cause firms to go from "not recruiting" to "recruiting," rather than increasing the count of recruits among employers who would have recruited anyway. As confirmation of this, taking the 6 percentage point increase on the extensive margin (taken from col. 1) and multiplying it by the control group conditional-on-positive number of ex ante recruits, which is 2.4, the expected increase in the count of recruits is 0.156, which is close to the realized increase of 0.151 (from the coefficient on the treatment indicator in col. 4).

B. Responses from Recruited Workers to ex ante Recruiting Invitations

A primary practical concern with algorithmically generated recommendations is their "quality," which we might think of as their usefulness to employers and workers alike. One proxy measure of recommendation quality is the discretion invited workers show when responding: workers would presumably respond more negatively to worse invitations (or invitations from worse employers). Using the sample of ex ante recruiting invitations, table 3 shows that there is no evidence that workers respond differently to recruiting invitations likely to have been experimentally induced, relative to

Table 3
Recruited Worker Responses to ex ante Recruiting Invitations

	Recruited Worker Response to Invitation		
	Responded? (1)	Accepted? (2)	Accepted? (3)
Treatment (Trt)	.0002	−.025	−.030
	(.019)	(.020)	(.026)
Technical (Tech)			−.049
			(.030)
Trt × Tech			.014
			(.040)
Constant	.588***	.454***	.474***
	(.014)	(.015)	(.019)
Observations	3,717	3,717	3,717
R^2	.00000	.001	.002

Note.—This table reports ordinary least squares regressions where the dependent variables are indicators for whether a recruited worker responded to an invitation (to either accept or decline), in col. 1, and whether they accepted the recruiting invitation and ultimately applied for the job opening, in col. 2. The sample consists of all ex ante recruited applicants to job openings assigned to the experiment where 10 or fewer ex ante recruiting invitations were sent. "Treatment" indicates that the job opening associated with the recruiting invitation was assigned to the treatment group. "Technical" indicates that the associated job opening required programming. Standard errors (in parentheses) are clustered at the level of the individual job opening.

*** $p \leq .001$.

the control group, suggesting that those experimentally induced invitations are similar to the "natural" invitations sent in the control group.

First, I examine whether the invited worker responded at all to the invitation, either to accept or reject. Column 1 of table 3 reports an estimate of

$$\text{Respond}_{ij} = \beta_0 + \beta_1 \text{Trt}_j + \varepsilon_j, \tag{12}$$

where Respond_{ij} is an indicator for whether the recruited worker i responded to the recruiting invitation to apply to job opening j, either yes or no ($\text{Respond}_{ij} = 0$ would mean that the recruited worker ignored the invitation). The sample is all ex ante recruiting invitations where the employer sent 10 or fewer ex ante recruiting invitations in total.[6] To account for the hierarchical nature of the data, standard errors are clustered at the level of the job opening. There is no evidence of systematic difference across experimental groups with respect to recruited worker responsiveness; the coefficient on the treatment indicator in column 1 is very close to zero, with a small standard error (slightly less than 2 percentage points).

Next, I look at whether the invited worker actually accepted the recruiting invitation. In columns 2 and 3, the dependent variable is whether the recruited worker accepted the invitation by submitting an application to the associated job opening. Column 2 shows that there is no evidence of a consequential treatment effect, as the coefficient is close to zero.[7] To test for differences by the nature of work, column 3 reports results from a specification that interacts the treatment indicator with an indicator for a technical job opening. Again, there is no evidence of a consequential difference across the experimental groups by the nature of work.

C. Size and Composition of the Applicant Pool

If the treatment increased ex ante recruiting and these recruited workers accepted employer recruiting invitations at the same rates across experimental groups, then the treatment should have increased the number of ex ante recruited applicants in the treated employer's applicant pool. Table 4 confirms this, showing that the treatment substantially increased the number of ex ante recruited applicants from which the employer had to choose.

Column 1 of table 4 reports a regression

$$R_j = \beta_0 + \beta_1 \text{Trt}_j + \varepsilon, \tag{13}$$

[6] I remove employers sending many invitations because they are more akin to spammers rather than bona fide employers. Further, by using an ordinary least squares model but then clustering standard errors to account for the hierarchical nature of the data, including "mass-invite" employers unduly weights their actions in the point estimates.

[7] Note that the baseline responsiveness and acceptance rates are close in magnitude, which means that few recruited applicants bother responding to say "no."

Table 4
Effect of the Treatment on the Number of Recruited Applicants

	Measures of ex ante Recruits in the Applicant Pool			
	Count		Any?	
	(1)	(2)	(3)	(4)
Treatment (Trt)	.076***	.076**	.043***	.032**
	(.016)	(.023)	(.008)	(.011)
Technical (Tech)		.002		−.004
		(.020)		(.010)
Tech × Trt		−.001		.021
		(.032)		(.016)
Constant	.138***	.137***	.091***	.093***
	(.010)	(.014)	(.005)	(.007)
Observations	6,127	6,127	6,127	6,127
R^2	.004	.004	.005	.005

NOTE.—This table reports ordinary least squares regressions where the dependent variable is either the count of recruited applicants, in cols. 1 and 2, or whether the job opening had any recruited applicants at all, in cols. 3 and 4. The sample is restricted to employers sending 10 or fewer ex ante recruiting invitations. "Treatment" is an indicator for whether the job opening was assigned the recommendations treatment. "Technical" is an indicator for whether the job opening required computer programming. Robust standard errors are reported in parentheses.
** $p \leq .01$.
*** $p \leq .001$.

where R_j is the count of recruited applicants to job opening j. For this analysis, I return to using the job openings experimental sample (rather than the recruited applicant sample used in table 3 in the previous section). The sample is restricted to employers sending 10 or fewer ex ante recruiting invitations. Column 1 shows that the treatment increased the number of recruits in the applicant pool by about 50%, raising the per opening count from about 0.14 to 0.21. To test whether treatment effects differed by job opening type, column 2 reports an alternative specification in which the treatment is interacted with the "technical" indicator. There is no evidence of heterogeneous treatment effects; the coefficient on the interaction term is close to zero, and the standard error is small relative to the baseline count of applicants.

In addition to measuring the count of recruiting applicants, I can also look at whether any recruited applicants applied at all. In column 3, the dependent variable is an indicator for whether any recruited applicants were in the applicant pool. There was approximately a 5 percentage point increase from the baseline of 10% in the control. Column 4 reports an alternate specification with the treatment indicator interacted with the job opening type; as before, there is no strong evidence of heterogeneous effects by the nature of work. Appendix Section A2.1 contains an analysis of the effect of the treatment on the total applicant pool size, and while the point estimates of the treatment effect are positive, they are highly imprecise. This imprecision is unsurprising given that the count of applicants has a high variance.

D. Characteristics of ex ante Recruited Applicants

The treatment increased the count of recruited applicants, but are these treatment-induced applicants comparable to the kinds of workers an employer would have recruited "on his/her own"? Table 5 shows that the experimentally induced recruits are statistically indistinguishable from the control group recruits on a number of dimensions employers are known to care about. It also shows that recruited applicants, regardless of source, are highly positively selected compared to nonrecruited organic applicants.

From an employer's perspective, perhaps the most consequential attribute of an applicant is his/her prior earnings at the time of application. Pallais (2014) shows that prior earnings are highly valued by employers. For this analysis, the sample consists of all job applications sent to job openings assigned to the experiment, where 100 or fewer applications were sent in total and the applicant had some amount of past earnings. Table 5, column 1 reports an estimate of

Table 5
Comparison of Applicant Characteristics by Recruiting Status and Treatment Assignment of the Applied to Job Opening

	Attributes of Applicants to the Job Opening			
	Log Prior Earnings (1)	Any Prior Earnings? (2)	Log Profile Rate (3)	Log Wage Bid (4)
Recruit (Recruit)	1.226***	.201***	.205*	.322***
	(.134)	(.014)	(.081)	(.081)
Treatment (Trt)	−.041	−.007	−.028	−.039
	(.041)	(.007)	(.026)	(.031)
Recruit × Trt	.061	−.004	.056	.044
	(.176)	(.021)	(.102)	(.101)
Constant	6.809***	.725***	2.042***	2.063***
	(.028)	(.005)	(.019)	(.022)
Observations	63,537	87,606	54,252	54,252
R^2	.006	.003	.002	.004

NOTE.—This table reports ordinary least squares regressions where the dependent variables are attributes of a job applicant or his/her job application at the time of application. The samples are constructed from all job applications to job openings assigned to the experiment. The sample in col. 1 is restricted to applicants with some prior earnings at the time of application. The col. 2 sample is the full set of applicants, regardless of past earnings. The same sample is used for the regressions in both col. 3 and col. 4. The sample consists of all job applications to hourly job openings where profile rates and wage bids were greater than \$1 but less than \$100. "Prior earnings" are total on-platform earnings, in US dollars, by the applicant before they applied. "Profile rate" is the hourly wage rate the worker lists on his/her public profile. "Wage bid" is the hourly wage bid the worker proposed when applying for the job opening. "Recruit" indicates that the applicant was recruited by the employer in the first hour after job posting (i.e., an ex ante recruit). "Treatment" indicates that the job application was to a job opening assigned to the treatment group. Standard errors (in parentheses) are clustered at the level of the individual job opening.

* $p \leq .05$.

*** $p \leq .001$.

$$\log y_{ij} = \beta_0 + \beta_1 \text{Recruit}_{ij} + \beta_2 \text{Trt}_j + \beta_3 \left(\text{Recruit}_{ij} \times \text{Trt}_j\right) + \varepsilon_j, \quad (14)$$

where Recruit_{ij} is an indicator for whether applicant i was a recruit to job opening j and Trt_j is the treatment indicator for opening j. The dependent variable y_{ij} is the cumulative prior earnings of an applicant when he applied to the job. Standard errors are clustered at the level of the job opening.

The coefficient on Recruit_{ij} in column 1 is large, positive, and economically important: in the control group, nonrecruited applicants have, on average, about \$900 in past earnings, whereas recruited applicants have slightly more than \$3,000 in past earnings. This same pattern holds in the treatment group, with recruited applicants being highly positively selected. There is no evidence that prior earnings for recruits differs by treatment assignment; the point estimate for the effect of the treatment corresponds to about 2 percentage points greater earnings, but this estimate is highly imprecise.

In addition to total earnings, employers might also care whether the applicant has any earnings at all. To test whether recruited applicants are positively selected on this dimension, column 2 reports a regression where the dependent variable is an indicator for whether the applicant had any past earnings. As with the amount of earnings, the data show that recruited applicants are far more likely to have past on-platform earnings (90% vs. a little more than 70%). Further, the same pattern holds in both the treatment and the control groups, with no significant differences by group.

Another measure employers care about is a worker's past wage rate. When a worker applies to an hourly job opening, his/her wage bid and profile rates are recorded. Columns 3 and 4 report regressions where the dependent variables are the log profile rate and log wage bid of the applicant, respectively. The samples for both estimates are restricted to applications sent to hourly job openings where both the profile rate and hourly charge rate were above \$1 but less than \$100. As with earnings, the large, positive coefficient on the Recruit indicator shows that recruits are positively selected: profile rates are about 25% higher (from col. 3 and wage bids are about 40% higher (from col. 4). The coefficient on the interaction term is positive and on the order of about 5 percentage points, though the estimates are imprecise and far from significant. If not a statistical artifact, the positive interaction term suggests that treatment-induced recruits were somewhat more expensive. Despite this possibility, at least within the limits of available statistical power, experimentally induced recruited applicants are similar to those in the control group in terms of past on-platform earnings (both amount and existence), wages bids, and profile rates.

E. Employer Screening, Measured at the Individual Applicant Level

Before employers make a hire, they evaluate, or "screen," their applicant pools. Not all applicants are screened; employers make choices about which applicants to screen on the basis of their observable characteristics. It is beyond the scope of this paper to model the employer's screening decision problem fully, but it is useful to see whether employers show any bias in their screening with respect to the treatment and the recruiting status of applicants. Table 6 shows that (i) employers are substantially more likely to screen recruited applicants, (ii) there is no evidence of screening "crowd-out" of nonrecruited applicants in the treatment (despite the treatment raising recruited applicant counts), and (iii) there is no strong evidence that screening probability differed for recruits by the treatment assignment of the applied-to job opening.

Table 6, column 1, reports an estimate of the regression

$$\text{Screened}_{ij} = \beta_0 + \beta_1(\text{Trt}_j \times \text{Recruit}_j) + \beta_2\text{Trt}_j + \beta_3\text{Recruit}_j + \varepsilon_j, \quad (15)$$

where Screened_{ij} is an indicator for whether the employer screened applicant i to job opening j. The sample consists of all job applications, and standard errors are clustered at the level of the job opening. From the intercept, it is apparent that the baseline screening rate is high but not nearly 100%—about 65% of organic applicants are screened. However, recruited applicants are at an advantage: their screening rate in the control group is slightly more than 9 percentage points higher. There is no strong evidence that screen-

Table 6
Screening Outcomes for Individual Applicants

	Employer Screening of Applicants in Their Pool: Viewed the Application?		
	(1)	(2)	(3)
Recruit (Recruit)	.093***	.056	.121**
	(.027)	(.034)	(.044)
Treatment (Trt)	−.013	−.017	−.006
	(.016)	(.020)	(.023)
Recruit × Trt	.011	.056	−.025
	(.036)	(.044)	(.056)
Constant	.650***	.700***	.605***
	(.011)	(.014)	(.016)
Opening type	All	Technical	Nontechnical
Observations	73,577	34,084	39,493
R^2	.001	.001	.001

NOTE.—This table reports ordinary least squares regressions where the dependent variable in each regression is whether the applicant was screened. The samples are constructed from all job applications to job openings assigned to the experiment. The sample in col. 1 includes all applications. The col. 2 sample is restricted to applicants to technical job openings, while the col. 3 sample is restricted to applicants to nontechnical job openings. "Recruit" indicates that the applicant was recruited by the employer in the first hour after job posting (i.e., an ex ante recruit). "Treatment" indicates that the job application was to a job opening assigned to the treatment group. Standard errors (in parentheses) are clustered at the level of the individual job opening.
** $p \le .01$.
*** $p \le .001$.

ing rates depend on the treatment assignment of the associated job opening for either recruits or organic applicants.

To test for differential screening by the nature of work, I estimate the column 1 regression for the two types of jobs separately. Column 2 reports the same applicant-level screening regression as in column 1 but for applications to technical job openings; column 3 reports results for applicants to nontechnical job openings. While in both samples recruits are more likely to be screened, the advantage is smaller for technical job openings. However, the column 2 intercept shows that the baseline probability of screening for technical job openings is higher by about 10 percentage points compared to nontechnical job openings, leaving less "room" for a large difference between technical and nontechnical job openings. As in column 1, there is no strong evidence that screening rates depend on treatment assignment, though the point estimate on the recruit indicator and treatment indicator interaction in column 2 is more than 5 percentage points. This would be an economically important effect, but it is not significant. If it is not a statistical artifact, it suggests that treatment-induced recruits for technical job openings were more likely to be screened.

F. Employer Hiring

The primary goal of offering algorithmic recommendations was to increase hiring. Table 7 shows that recommendations were effective at increasing hires, but only unambiguously so for technical job openings. The differential treatment effectiveness does "show up" in the overall analysis

Table 7
Effects of the Recommendations Treatment on Hiring

	Dependent Variable			
	Hire Made		Hired Early Recruit	Hired Organic
	All (1)	Technical (2)	Technical (3)	Technical (4)
Treatment	.019	.043***	.020***	.022
	(.011)	(.015)	(.007)	(.015)
Constant	.276***	.216***	.032***	.212***
	(.008)	(.010)	(.004)	(.010)
Observations	6,209	3,136	3,136	3,136
R^2	.0004	.003	.002	.001

Note.—This table reports several regressions where the outcome variables are measures of employer hiring, using data from the recommendations experiment. Each regression was estimated using ordinary least squares, and robust standard errors (in parentheses) are reported. The key independent variable across across regressions is the indicator for whether the vacancy was assigned to the treatment group. The dependent variable in each of these regressions is whether or not the employer hired a worker of a particular type. In cols. 1 and 2, the indicator is for hiring anyone at all. In col. 1 the sample is all vacancies, while in col. 2 the sample consists of only technical vacancies. In col. 3, the outcome is whether the employer hired a recruited applicant, while in col. 4 the outcome is whether the employer hired an organic applicant. In both col. 3 and col. 4, the sample is restricted to technical vacancies.

*** $p \leq .001$.

in that treated job openings were more likely to surpass \$500 in total wages paid, which is close to the median spend per job. This finding is consistent with the treatment increasing formation of relatively technical job matches, which tend to have higher wage bills. Hypotheses for why treatment effects differed by the type of work will be explored in subsequent subsections, but first I simply present the results. Among technical job openings, I find no evidence of crowd-out of organic applicants, but I show that given the size of the point estimates of the treatment, it would be unlikely to find strong crowd-out.

Column 1, table 7, reports an estimate of

$$\text{EmpHired}_j = \beta_0 + \beta_1 \text{Trt}_j + \varepsilon, \tag{16}$$

where EmpHired_j is an indicator for whether some amount of money was spent by the employer on a worker hired for job opening j. The coefficient on the treatment indicator is positive, with a magnitude of nearly 2 percentage points, but this estimate is not conventionally significant. Column 2 reports the same regression as column 1, but with the sample restricted to technical job openings. Among technical job openings, the treatment increased the probability that a hire was made by more than 4 percentage points, from a base of only 22%. This subgroup effect is easily still significant under the conservative Bonferroni correction for multiple hypotheses testing.

The increase in the fill rate among technical job openings shown by column 2—a 4.3 percentage point increase—is a large effect, too large in fact to be credibly explained solely by the direct channel of increased ex ante recruiting. Recall that the treatment increased recruiting by 7.3 percentage points among technical job openings. Technical job openings in the control group had a baseline recruiting rate of 15% and a fill-from-recruits rate of 3.8%. If one assumes that the marginal increase in recruits from the treatment "converted" to hires at the same rate in the treatment as in the control, the treatment effect in column 2 should be $(1/4)0.07 \approx 1.8\%$. This is close to the treatment effect of 2 percentage points found in the column 2 regression, where the dependent variable is whether the employer hired an ex ante recruit.

The "extra" increase in hiring find in column 2 comes from the firm hiring organic applicants: column 3 reports a regression where the dependent variable is the firm hiring an organic applicant, and as expected, the coefficient on the treatment indicator is positive and has the right magnitude to explain the overall increase in hiring. Although the coefficient on the treatment indicator in column 4 is not conventionally significant, recall from the employer recruiting model that the treatment effect should be negative because recruits should crowd out nonrecruited applicants.

The lack of crowd-out—and perhaps even complementarity—may seem surprising, but not if one considers the overall low fill rate. Crowd-out would occur among employers who would have hired an organic applicant when in the control but would instead hire a recruited applicant when in the treatment. Column 3 shows that only 20% of employers in the control group hired an organic applicant, 3% hired a recruit, and the rest did not hire anyone. If the approximately 7 percentage point increase in recruiting caused by the treatment was uniform across all of these segments, then one would see a $(0.07)(0.20) \approx 1.4\%$ increase in recruiting among organic-hiring control employers if they had counter-factually received the treatment. Under the assumption that the conversion rate of recruits into hires is the same value, that is, ~1/4, and that every one of these hired recruits displaced an organic applicant, the reduction in organic hiring would only be $(1/4)(0.014) \approx -0.0035$, which is a very small effect. Of course, if the fill rate were higher—or if the treatment was heterogeneous—then crowd-out could become substantial.

G. Hiring at the Applicant Level

In addition to measuring hiring at the level of the job opening, it is useful to consider hiring at the level of the individual applicant. The applicant-level view allows me to test whether employers show a preference for recruited applicants and whether their preferences depended on the treatment assignment and the nature of work. Table 8 shows that recruited applicants are far more likely to be hired in general for both technical and nontechnical categories of work. This preference partially explains why the treatment can have such a strong effect on hiring despite the number of induced applicants from the treatment being small relative to the size of the applicant pool. As one would expect given that the treatment increased the number of recruited applicants but not the fraction of employers hiring recruits for that category, there is some evidence that treatment-induced applicants in nontechnical categories of work have a lower hire rate than their control counterparts. A potential reason for this difference is explored in the next section in which I compare recruited applicants across experimental groups and categories of work on the basis of their observable characteristics.

Column 1 of table 8 reports an estimate of

$$\text{Hired}_{ij} = \beta_0 + \beta_1 \text{Recruit}_{ij} + \beta_2 \text{Trt}_j + \beta_3 \left(\text{Recruit}_{ij} \times \text{Trt}_j\right) + \varepsilon_j, \quad (17)$$

where the sample is all applicants to a job opening and the Hired_{ij} is an indicator for whether worker i was hired for job opening j. Standard errors are clustered at the level of the individual job opening. The constant term from column 1 shows that the baseline hire rate for organic applicants in the control is about 1.5%, but for recruits, this goes up by slightly more than 10 percentage points. There is no evidence that these hire rates differ substantially by treatment assignment.

Table 8
Hiring Outcomes by Treatment Assignment Job Opening Type and Recruit Status

	Dependent Variable: Applicant Hired?		
	(1)	(2)	(3)
Recruit (Recruit)	.103***	.083***	.129***
	(.015)	(.018)	(.024)
Treatment (Trt)	.001	.003*	−.001
	(.001)	(.001)	(.001)
Recruit × Trt	−.010	.022	−.047
	(.019)	(.026)	(.028)
Constant	.015***	.015***	.016***
	(.001)	(.001)	(.001)
SEs clustered at job opening?	Yes	Yes	Yes
Include technical?	Yes	Yes	No
Include nontechnical?	Yes	No	Yes
Observations	73,577	34,084	39,493
R^2	.011	.011	.011

Note.—This table reports ordinary least squares regressions where the dependent variable in each regression is whether the applicant was hired. The samples are constructed from all job applications to job openings assigned to the experiment. The sample in col. 1 includes all applications. The col. 2 sample is restricted to applicants to technical job openings, while the col. 3 sample is restricted to applicants to nontechnical job openings. "Recruit" indicates that the applicant was recruited by the employer in the first hour after job-posting (i.e., an ex ante recruit). "Treatment" indicates that the job application was to a job opening assigned to the treatment group. Standard errors (in parentheses) are clustered at the level of the individual job opening.
* $p \leq .05$.
*** $p \leq .001$.

In column 2, the sample is restricted to job applicants to technical job openings. The hire rate for recruits in the control is about 2 percentage points lower than the hire rate in the control from column 1, but a formal hypothesis test would fail to reject a null of no difference in the point estimates. Where there is a difference is in the hire rate of organic applicants in the treatment—the coefficient on the treatment indicator is about 3/10ths of a percentage point. Given the finding of a higher rate of organic hires in the treatment group for technical categories, this result is, in a sense, mechanically expected. However, the greater statistical precision offered by the job applicant view of table 8 and the resultant stronger finding raises the question of whether the increased hiring of organic applicants in the treatment, for technical job openings, is not simply a statistical artifact. One interesting possibility is that the existence of recruited applicants increased employer screening on either the extensive or intensive margins, which in turn "spilled over" onto nonrecruits, though this is highly speculative.[8]

[8] In a re-analysis of several well-known correspondence studies, Phillips (2015) argues that designs that sent multiple applications to the same employer induced exogenous variation in applicant pool quality, which in turn generated search externalities.

To investigate the technical/nontechnical distinction further, in column 3 the sample is restricted to nontechnical job openings. As in the other regressions, there is a large positive coefficient on the recruit indicator, showing a strong preference among control employers for hiring recruited applicants for nontechnical job openings. The coefficient is larger than in column 1, but the standard errors are large enough that a formal hypothesis test would reject a difference in hiring preference for recruits across samples. Unlike technical job openings, there is no evidence that nontechnical organic applicants benefited from the treatment—the coefficient on the treatment indicator is very close to zero and precisely estimated. The coefficient on the recruit indicator and treatment indicator interaction term is large in magnitude and negative, as expected given that the treatment increased recruiting but not hiring. The possibility that treatment-induced recruits for nontechnical job openings systematically differed is the focus of the next section.

H. Comparison of ex ante Recruits by Job Opening Type and Associated Treatment Assignment

One potential explanation for why the treatment was only effective for technical job openings was that the algorithm delivered relatively poorer recommendations for nontechnical job openings. Table 9 compares recruits across experimental groups and categories of work to look for differences. Within the limits of the statistical power available, there are no statistically significant differences in ex ante recruits across experimental groups, conditioned on the job type. However, there is some evidence that the recruits induced by the treatment were relatively more experienced and thus more expensive. If experimentally induced nontechnical recruits were in fact relatively more expensive, it could explain why they have a lower hire rate. A difference in price could have a magnified effect in nontechnical categories, as Horton and Johari (2015) show that employers in nontechnical categories of work are far more price sensitive than those hiring in technical categories of work. While higher relative cost is an attractive explanation, the general imprecision of these estimates make it difficult to reach strong conclusions.

Column 1, table 9, reports an estimate

$$\log y_{ij} = \beta_0 + \beta_1 \mathrm{Trt}_{ij} + \beta_2 \mathrm{Technical}_j + \beta_3 \left(\mathrm{Technical}_{ij} \times \mathrm{Trt}_j\right) + \varepsilon_j, \quad (18)$$

where the sample is restricted to ex ante recruits and y_{ij} is the past cumulative earnings of applicant i to job opening j. Standard errors are clustered at the level of the job opening. As expected, the coefficient on the technical indicator is positive and very large: recruits to technical job openings have nearly 70% higher earnings. The coefficient on the treatment indicator is small—implying little more than a 1 percentage point increase in earnings for recruits to nontechnical job openings—but the estimate is highly imprecise. The technical and treatment interaction is large as a point estimate, implying

Table 9
Comparison of Recruited Applicant Characteristics, by Treatment Assignment

	Dependent Variable			
	Log Prior Earnings (1)	Any Prior Earnings? (2)	Log Profile Rate (3)	Log Wage Bid (4)
Technical (Tech)	.694**	.012	.667***	.690***
	(.256)	(.028)	(.135)	(.133)
Treatment (Trt)	.049	.004	.127	.104
	(.227)	(.029)	(.124)	(.124)
Tech × Trt	.168	−.032	−.067	−.066
	(.323)	(.041)	(.176)	(.171)
Constant	7.643***	.920***	1.845***	1.970***
	(.174)	(.021)	(.094)	(.096)
SEs clustered at job opening?	Yes	Yes	Yes	Yes
Observations	1,340	1,460	961	961
R^2	.035	.001	.141	.153

Note.—This table reports ordinary least squares regressions where the dependent variables are attributes of an ex ante recruited job applicant or his/her job application, at the time of application. The samples are constructed from all job applications by ex ante recruits to job openings assigned to the experiment. The sample in col. 1 is restricted to applicants with some prior earnings at the time of application. The col. 2 sample is the full set of applicants, regardless of past earnings. The sample used for both col. 3 and col. 4 is the same. It consists of all job applications to hourly job openings where profile rates and wage bids were greater than $1 but less than $100. "Prior earnings" are total on-platform earnings, in US dollars, by the applicant before they applied. "Profile rate" is the hourly wage rate the worker lists on his/her public profile. "Wage bid" is the hourly wage bid the worker proposed when applying for the job opening. "Recruit" indicates that the applicant was recruited by the employer in the first hour after job posting (i.e., an ex ante recruit). "Treatment" indicates that the job application was to a job opening assigned to the treatment group. Standard errors are clustered at the level of the individual job opening.
** $p \le .01$.
*** $p \le .001$.

a 17 percentage point increase in the treatment, but again, the estimate is highly imprecise and far from conventionally significant.

In column 2, the dependent variable is an indicator for any prior earnings. As compared to prior earnings, precision is better and the differences by group and job type are small—though given that over 90% of recruits have some prior experience, there is little room for large differences. On profile rates (in col. 3) and wage bids (col. 4), it is apparent that technical recruits have far higher profile rates and submit higher bids. Note that for these regressions the sample is limited to hourly job openings.

Also mirroring the pattern from column 1, the standard errors are large enough to contain economically important effects comfortably. For example, the coefficient on the treatment indicator implies that wage bids and profile rates were about 10% higher for recruits to nontechnical job openings in the treatment than in the control. These results are not conventionally significant—far from it in fact—but if the point estimates are correct, it would imply that the algorithmic recommendations were inducing relatively more expensive applicants for nontechnical openings. The coefficient on the technical and treatment interaction term is negative, implying that experimentally induced technical recruits were relatively cheaper—but again this effect is far from significant.

I. Treatment Effects on Hiring, Conditioned by Predicted Organic Applicant Count

In the model presented in Section IV.A, a reduction in recruiting costs increases the number of recruits for all employers, but the effects on fill rates are larger for employers with smaller expected applicant pools. This aspect of the model potentially provides a parsimonious explanation for the technical/nontechnical difference in treatment effectiveness. Because employers with technical job openings have smaller applicant pools, they value the pool-expanding effects of the treatment more than employers with nontechnical job openings.

It is tempting to test this hypothesis by conditioning the treatment indicator on the realized count of organic applicants and test for heterogeneous effects. However, using the realized organic application count as a regressor is problematic. First, idiosyncratic job opening attributes might affect both the count of applicants received and the baseline probability of a hire being made. Second, the count of organic applications could be affected by the treatment—if, for example, the employer makes a quick hire because he/she has access to the treatment. This possibility makes the organic applicant count an inappropriate right-hand-side variable.

Rather than use the realized application count, I instead use the predicted organic application count, with predictions derived from a machine learning model fit only with pre-randomization job opening attributes—see Appendix A.3 for details on the predictive model. I estimate the effects of the treatment on hiring, conditional upon predicted organic application counts. Table 10 shows that the returns from the treatment depend strongly upon the predicted organic applicant count, with effectiveness decreasing in the predicted applicant pool size. This dependency offers a parsimonious explanation for the difference in treatment effects on hiring for technical and nontechnical job openings.

Table 10, column 1, reports an estimate of

$$\text{Hired}_j = \beta_0 + \beta_1 \text{Trt}_j + \beta_2 \widehat{\log A_j} + \beta_3 \left(\text{Trt}_j \times \widehat{\log A_j}\right) + \varepsilon_j, \qquad (19)$$

where Hired_j is an indicator for whether a hire was made against that job opening and $\widehat{\log A_j}$ is the log predicted count of organic applicants for opening j. The sample for column 1 is all job openings assigned to the experiment. The coefficient on the treatment indicator is positive and large: when the predicted organic applicant count is one (as $\log(1) = 0$), the model implies the treatment causes an 8 percentage point increase in hiring. Recall that estimated treatment effect on hiring for the full sample was about 2 per-

Table 10
Effects of the Recommendations Treatment, Conditioned upon Predicted Organic Applicant Counts

	Dependent Variable: Hire Made?		
	(1)	(2)	(3)
Treatment (Trt)	.081*	.104+	.044
	(.040)	(.059)	(.054)
Predicted log number of organic applicants ($\widehat{\log A}$)	.053***	.011	.072***
	(.012)	(.019)	(.016)
Trt × $\widehat{\log A}$	−.029	−.029	−.024
	(.018)	(.027)	(.024)
Constant	.161***	.193***	.180***
	(.028)	(.041)	(.037)
Include technical?	Yes	Yes	No
Include nontechnical?	Yes	No	Yes
Observations	6,209	3,136	3,073
R^2	.004	.003	.009

NOTE.—This table reports ordinary least squares regressions where the dependent variable is whether the employer made a hire. In col. 1, the entire experimental sample is used, while in cols. 2 and 3, the sample consists of all technical and nontechnical job openings, respectively. In each regression, the treatment indicator is interacted with the prediction of how many organic applicants the job opening would receive. See Appendix Sec. A3 for details on the predictive model.
+ $p \leq .10$.
* $p \leq .05$.
*** $p \leq .001$.

centage points (and not conventionally significant). This is not a wildly out-of-sample extrapolation, as a substantial fraction of job openings receive just a small number of organic applicants. The coefficient on the treatment and predicted applicant count interaction term is negative, implying that as the count of organic applicants increases, the treatment is less effective at increasing hiring. In effect, the treatment and organic applicants are substitutes. If one takes the linear model seriously, the treatment no longer has a positive effect at $\exp(0.80/(0.054 - 0.028)) \approx 22$ applicants. However, there are good reasons to not take it too seriously—in columns 2 and 3, I estimate the model separately for technical and nontechnical jobs and find important qualitative differences by category of work.

Column 2 is the same regression as column 1, but with the sample restricted to technical job openings. Compared to the column 2 regression, both the coefficient on the treatment indicator and the interaction term are similar in magnitude. However, the returns to predicted organic applicants are only a fifth of what they were in the full estimate from column 1. Of course, the count of predicted organic applicants is not as good as randomly assigned, and as such, the coefficient cannot be interpreted causally. However, much of the variation in the predicted applicant count is likely due to supply factors—the subcategory of work is by far the most important component of the predictive model. To the extent that there is exogenous variation in organic applicant counts, the evidence suggests that technical job openings get little value from the marginal organic applicant. This is consistent with Horton and Johari (2015) in that employers with technical job categories are much less price sensitive and thus would have a relative preference for the higher-quality but higher-wage applicants that come through the employer recruiting channel.

Column 3 is the same regression as column 1, but with the sample restricted to nontechnical job openings. The coefficient on the treatment indicator is positive—recall that it was negative without conditioning on the predicted organic applicant count. However, it is substantially smaller than the effect estimates from column 2, which was the technical-only regression. The interaction term is negative and similar in magnitude to the other estimates, suggesting that as with technical jobs, organic applicants and recruits are substitutes. The big qualitative difference in column 3 versus the other regressions is in the returns to more organic applicants: as the organic application count increases, hiring goes up substantially, with an effect about six times larger than the estimate from the technical job openings.

VI. Discussion

The main results of the experiment are clear: (i) employers act upon recommendations, with the effect mainly being on the extensive margin; (ii) recommendations deliver additional applicants similar to the kinds of workers that employers recruit on their own; (iii) the recommendations were effective at increasing hiring in technical job openings, without any detectable crowd-out of non-recruited organic applicants.

Explaining (i) and (ii) is fairly straightforward: the treatment lowered the cost of recruiting, and more employers decided to recruit. The comparison of ex ante recruited applicants and organic applicants shows that recruits—in both the treatment and control groups—are highly positively selected. By delivering more applicants that "look like" the kinds of applicants employers have a strong preference for hiring, hiring increased at least for some types of openings. Before discussing the technical/nontechnical treatment effect heterogeneity, it is useful to consider explanations for treatment efficacy and compare them to other facts uncovered by the applicant-level analyses.

There is no evidence that the main effect of the experiment was simply to alert employers to the possibility of recruiting, as the treatment did not increase "late" recruiting. The treatment-induced recruited workers who applied were similar on observables to the kinds of workers employers tend to recruit, and employers screened and hired these recruits at similar rates—undercutting the notion that employers regarded the treatment-induced recruits as better or worse than their observables would suggest. Furthermore, there is some "downstream" evidence of comparability of experimentally induced recruits to control groups: if the treatment caused employers to believe that their pools were more productive than the control, even conditioning on observables, then they would, on average, be relatively disappointed by realized performance, and yet there is no evidence of this—treatment and control matches look very similar with respect to all measures of match quality. (Match quality analysis is in Appendix Section A2.3).

The largest puzzle in the experimental results is why the treatment only increased hiring for technical job openings. There is no strong evidence that the recommendations were any worse for nontechnical job openings: recruits were similar on observables, and "uptake" by employers and workers was the same by type of work and by experimental group. If there is a difference, it is that nontechnical treatment recruits tended to be slightly more expensive than their control counterparts. The explanation most consistent with all the facts is that technical job openings differed from nontechnical job openings in two important ways. First, technical job openings generally get fewer applicants and value the marginal applicant more. Second, employers with technical job openings are less price sensitive and value higher-quality applicants relatively more than employers with nontechnical job openings. As the algorithmic recommendations deliver more applicants—particularly higher-quality and higher-cost applicants—the treatment is more useful in technical categories of work.

VII. Conclusion

In this paper, I demonstrate that algorithmic recommendations are both acted upon by employers and also effective at raising hiring, at least for some kinds of job openings where more applicants of high quality are valued. While the algorithm is a "black box," it delivers recommendations observationally identical to the kinds of workers employers recruit in the absence of these recommendations, at least within the limits of available measurements and statistical power. As such, it makes sense to think of algorithmic recommendations as substituting for costly employer effort.

A novel feature of this experiment was that it was focused on helping the demand side of the labor market; most other active labor market policies have focused on the supply side. Perhaps the encouraging results from this study and its demand side focus are a coincidence, but an intriguing possibility is that serving firms is more fruitful than serving workers. There is a superficial symmetry between job openings and workers; job openings can be readily created and destroyed by employers at will, and while workers do enter and exit the labor market, it seems likely that the employer decision to create and fill an opening is more elastic with respect to assistance than the labor force participation of an individual worker. Turning to the conventional market analog of this experiment, for-profit recruiting firms offer their services primarily to companies rather than individuals.

In terms of generalizability, these results come from a context where search frictions are already very low, suggesting that bigger gains are possible in traditional, more friction-prone markets. As more of the labor market becomes computer-mediated, the possibilities for platform-based interventions grow in scope and power. Platforms invariably collect enormous amounts of data on market behaviors and outcomes; they also have nearly full control over what information market participants can see, and when. This possibility could have enormous equity and efficiency consequences for labor markets, and more studies on how to use this power constructively are needed.

Appendix

A1. Internal Validity

A1.1. Randomization

Job openings were randomly assigned to the treatment or control group after being posted to the marketplace. As such, pre-assignment covariates describing vacancy should be balanced across the groups; the count of observations across groups should also be consistent with a random process. In table A1, I confirm that this is the case. In the top panel, I report the count of observations and the p-value for a chi-square test. I cannot reject the null hypothesis of an identical assignment probability across the groups. In the lower panel, I report the fraction of vacancies of different "types" (by category of work) by experimental group. I also report two noncategory covariates—an indicator for whether the length of job description exceeded the median and whether the employer required that applicants have prior on-platform work experience. For each covariate, I report the mean and standard error for each group; I also report the difference in means, the standard error for that difference and the p-value for a two-sided t-test. I find good balance across the collection of covariates chosen, with none of the t-test p-values being conventionally significant.

A1.2. SUTVA

One concern with a field experiment is that the treatment cells interact with each other in some way, violating the stable unit treatment value assumption (SUTVA) required for causal inference. Any sufficiently large experiment conducted in a marketplace risks SUTVA problems. In this experimental setting, "deep" SUTVA concerns (say, moving the market or changing the incentive to entry or exit) are probably unwarranted. First, this experiment was conducted a year after Pallais (2014), who did find evidence of market-moving effects, and the market was at least twice as large when this experiment was run. Further, unlike Pallais's experiment, vacancies associated with all types of work were eligible. And because only new employers were eligible for the experiment and only some of these were allocated, only about 6% of the vacancies posted during the experimental period were actually assigned to the experiment (and only half of those were in the active treatment).

A2. Auxiliary Analyses

A2.1. Job Opening Applicant Counts

Although the treatment increased the number of recruited applicants, did it increase the size of the applicant pool in total? While the data are consistent with this hypothesis—treated employers had larger applicant pools on average—this difference in mean counts is very imprecisely estimated and far from conventionally significant. Table A2 shows that even with transformations designed to reduce the standard error, treatment estimates are quite imprecise.

In column 1, the coefficient on the treatment indicator is positive, but it is imprecisely estimated. The high variance in the per-opening number of applicants makes it difficult to detect any treatment effects. In column 2, I winsorize the per opening number of applicants with a ceiling of 50. This shrinks the standard error, but it also reduces the effect size, implying that some treatment openings had very high application counts.

A2.2. Match Attributes—Total Spending

Detecting differences in match outcomes across the experiment is challenging because the marginally filled job opening is not separately identifiable—it is pooled with all other filled job openings. A second complication is picking reasonable measures of match quality. One plausible economic measure of match quality is the total amount spent. If one assumes the market is competitive but that there is the possibility of match-specific surplus, then higher spending implies the employer wanted to buy more of what the worker was selling. Table A3 shows that although spending was higher among treated employers and that the effect sizes would be economically consequential, none of the effects are conventionally significant. There is no strong evidence that treatment-induced matches were any better or worse than those in the control, but the main conclusion is that the study is simply under-powered to detect even consequential differences in match quality.

Column 1 of table A3 reports an estimate of

$$\log Y_j = \beta_0 + \beta_1 \mathrm{Trt}_j + \varepsilon, \tag{A1}$$

where Y_j is the total spend of employer j on their job opening. The sample is restricted to employers with total spend greater than one dollar, that is, $Y_j \geq \$1$. The coefficient on the treatment indicator is positive and the magnitude is economically significant, with treated job openings having nearly 14 percentage point higher total spend. However, this estimate is far from conventionally significant.

The treatment only increased hiring in the technical categories of work. Column 2 reports an estimate of the same specification as in column 1, but with the sample restricted to technical job openings. The coefficient on the treatment indicator is slightly larger than in column 1—about 2 percentage points—but the estimate is even less precise, as expected given the smaller sample size. In column 3, the sample is restricted to nontechnical job openings. And although the coefficient on the treatment indicator is positive, the magnitude is smaller than in either column 1 or column 2, and the estimate is far from conventionally significant. Collectively, although there is some evidence of greater spend in treated job openings that are filled, the study is simply underpowered to detect even economically consequential changes in spend.

One potential concern with the finding that the treatment increases hiring for technical job openings is that perhaps these induced matches were small and inconsequential. Using a transformation of total spend as an outcome, this hypothesis is testable. In column 4, the dependent variable is whether the total spend was more than \$500, with the sample restricted to technical job openings (but including even unfilled job openings in the sample). The effect is large—about 2 percentage points off the base of only about 7%, and it is conventionally significant.

A2.3. Match Attributes—Subjective Measures

When an employer on oDesk dissolves a relationship, he/she may give a reason for ending the contract. One listed reason the employer can select is the "job was completed successfully," which is an attractive proxy for match quality. Another qualitative measure is the feedback employers and workers give to each other at the conclusion of a contract. Table A4 shows that on these subjective measures, there is no strong evidence that the treatment assignment or the category of work mattered, though as with spend as a measure of match outcome, the study is underpowered to detect even consequential effects.

Column 1 of table A4 reports an estimate of

$$\text{Success}_j = \beta_0 + \beta_1 \text{Trt}_{ij} + \beta_2 \text{Technical}_j + \beta_3 \left(\text{Technical}_{ij} \times \text{Trt}_j\right) + \varepsilon, \quad \text{(A2)}$$

where Success_j is an indicator for whether the employer rated his/her completed job opening j a success. The intercept from the column 1 regression shows that slightly more than 70% of completed jobs are rated as having been completed successfully. Neither the treatment assignment nor the category of work had a detectable effect on the fraction of employers rating the project as being completed successfully. In column 2, the outcome variable is the worker's feedback (1–5 stars) on the employer, and in col. 3, the outcome variable the employer's feedback on the worker. This feedback is not required and so is not universally available for matches. In both cases, neither the treatment group nor the job opening type had a detectable effect on the ratings.

A3. Predictive Model of Organic Applicant Counts

To provide a measure of expected organic applicant counts, I fit a linear model, using the lasso (Tibshirani 1996) for feature selection and to prevent overfitting. I used cross-validation (Friedman, Hastie, and Tibshirani 2009) to select the optimal regularization parameter. The sample included all observations from the experiment where the job opening had at least one organic applicant.

The outcome was the log organic applicant count. The candidate predictors were the full set of dummies for the subcategory of the job opening (employers classify their job opening into one of 74 mutually exclusive categories of work), the length of the job description, dummies for the the employer's visibility setting (employers can restrict who can see their job openings), a dummy for whether the employer required applicants to have any past on-platform experience, a dummy for whether the job was hourly (compared to fixed price), and a dummy for whether the job description had an attachment. I also included all of the interactions of these predictors with themselves.

Table A1
Pre-treatment Covariate Means by Experimental Group

	Treatment Mean: $\bar{X}_{\text{Trt}}$	Control Mean: $\bar{X}_{\text{Ctl}}$	Difference in Means: $\bar{X}_{\text{Trt}} - \bar{X}_{\text{Ctl}}$	p-Value
Observation count	3,027	3,182		1.7
Type of work:				
Technical (1 if yes, 0 otherwise)	.500	.510	−.011	
	(.009)	(.009)	(.013)	.392
Nontechnical	.500	.490	.011	
	(.009)	(.009)	(.013)	.392
Type of work (more detailed):				
Administrative	.081	.081	−.000	
	(.005)	(.005)	(.007)	.947
Writing	.126	.122	.004	
	(.006)	(.006)	(.008)	.611
Web	.375	.389	−.014	
	(.009)	(.009)	(.012)	.243
Design	.114	.115	−.001	
	(.006)	(.006)	(.008)	.897
Software	.125	.121	.004	
	(.006)	(.006)	(.008)	.670
Vacancy attribute:				
Job description length > median	.562	.572	−.010	
	(.009)	(.009)	(.013)	.441
Required prior oDesk experience	.120	.109	.011	
	(.006)	(.006)	(.008)	.179

NOTE.—This table reports data from a recommendation experiment conducted on the oDesk platform. In the experiment, the treated group of employers received algorithmically generated recommendations of candidates for their vacancies, while the control group did not. The unit of randomization was the posted vacancy. The top panel, labeled "Observation count," is the number of vacancies per experimental group. The "Observation count" panel p-value is for a two-sided test of equal group assignment probabilities; the p-value was calculated by simulation, using the fact that 13,259 potential employers were allocated, but only 5,953 generated usable vacancies for assignment. In the lower panels, for each variable, we report the mean and standard error (in parentheses) for various pre-treatment covariates, as well as the standard error for the cross-group differences. The p-value column is for a two-sided t-test against the null hypothesis of no difference in means across the treatment and control groups. In the middle panels ("Type of work" and "Type of work—[more detailed]"), the covariantes are indicators for whether the vacancy was for a particular type of work. In the bottom panel, the two covariates are (i) an indicator for whether the employer required that applicants have prior oDesk experience and (ii) whether the count of text characters in the job description was greater than the median count for the pooled sample of all job descriptions in the experiment.

Table A2
Applicant Pool Size by Treatment

	Dependent Variable			
	Number of Applications	Number of Applications (Winsorized)		Any Applications?
	(1)	(2)	(3)	(4)
Treatment (Trt)	.805	.129	.372	.008
	(.693)	(.360)	(.571)	(.008)
Technical (Tech)			−2.593***	
			(.499)	
Trt × Tech			−.543	
			(.718)	
Constant	16.505***	14.371***	15.694***	.896***
	(.407)	(.249)	(.397)	(.005)
Observations	6,209	6,209	6,209	6,209
R^2	.0002	.00002	.010	.0002

NOTE.—This table reports several ordinary least squares regressions where the dependent variables are measures of size of the employer's total applicant pool. In col. 1, the dependent variable is the raw count of applicants, while in col. 2 the applicant count is winsorized at 50. Column 3 has the same dependent variable as in col. 2, but the Treatment indicator is interacted with the Technical indicator. The dependent variable in col. 4 is whether the employer received any applications at all. Robust standard errors are reported in parentheses.
*** $p \leq .001$.

Table A3
Effects of the Recommendations Treatment on the Wage Bill by Category of Work

	Dependent Variable			
	Log Total Spend			Total Spend > $500?
	(1)	(2)	(3)	(4)
Treatment	.138	.155	.051	.019**
	(.086)	(.127)	(.109)	(.009)
Constant	4.813***	5.420***	4.408***	.066***
	(.061)	(.092)	(.076)	(.007)
Include technical?	Yes	Yes	No	Yes
Include nontechnical?	Yes	No	Yes	No
At least $1 in spend?	Yes	Yes	Yes	No
Observations	1,759	740	1,019	3,136
R^2	.001	.002	.0002	.001

NOTE.—This table reports ordinary least squares regressions where the outcome variables are measures of employer spending. In cols. 1–3, the outcome is the log total spend, conditional upon at least $1 being spent. In column 4, the dependent variable is indicator for whether the employer spent more than $500, with the sample restricted to technical job openings. Robust standard errors are reported in parentheses.
** $p \leq .01$.
*** $p \leq .001$.

Table A4
Effect of Treatment on Qualitative Match Outcomes Conditional upon Outcome Being Available

	Dependent Variable		
	Employer Rated a Success? (1)	Worker-on-Employer Feedback (2)	Employer-on-Worker Feedback (3)
Treatment	.021	−.032	.074
	(.028)	(.061)	(.071)
Technical	−.025	.006	.083
	(.031)	(.066)	(.079)
Treatment × technical	−.019	.077	−.099
	(.044)	(.094)	(.111)
Constant	.714***	4.668***	4.345***
	(.020)	(.043)	(.049)
Observations	1,770	1,541	1,436
R^2	.002	.001	.001

Note.—The table reports ordinary least squares regressions where the dependent variable is some employer- or worker-reported measure of match quality. In col. 1, the dependent variable is whether the employer rated the project a success when ending the contact. In col. 2, the dependent variable is the rating (1–5 stars) by the worker on the employer. In col. 3, the dependent variable is the rating (1–5 stars) by the employer on the worker. For all three regressions, the samples consist of filled job openings where the associated outcome measure is available. Parties are not universally compelled to give these reports, so they are missing for some filled and ended job openings. Robust standard errors are in parentheses.
*** $p \leq .001$.

References

Adomavicius, Gediminas, and Alexander Tuzhilin. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering* 17, no. 6:734–49.

Agrawal, Ajay K., John Horton, Nico Lacetera, and Elizabeth Lyons. 2015. Digitization and the contract labor market: A research agenda. In *Economic analysis of the digital economy*, ed. Shane Greenstein, Avi Goldfarb, and Catherine Tucker. Chicago: University of Chicago Press.

Agrawal, Ajay K., Nicola Lacetera, and Elizabeth Lyons. 2013. Does information help or hinder job applicants from less developed countries in online markets? NBER Working Paper no. 18720 (January), National Bureau of Economic Research, Cambridge, MA.

Autor, David H., and David Scarborough. 2008. Does job testing harm minority workers? Evidence from retail establishments. *Quarterly Journal of Economics* 123, no. 1:219–77.

Barron, John M., and John Bishop. 1985. Extensive search, intensive search, and hiring costs: New evidence on employer hiring activity. *Economic Inquiry* 23, no. 3:363–82.

Barron, John M., Dan A. Black, and Mark A. Loewenstein. 1989. Job matching and on-the-job training. *Journal of Labor Economics* 7, no. 1: 1–19.

Burdett, Kenneth, and Elizabeth J. Cunningham. 1998. Toward a theory of vacancies. *Journal of Labor Economics* 16, no. 3:445–78.
Card, David, Jochen Kluve, and Andrea Weber. 2010. Active labour market policy evaluations: A meta-analysis. *Economic Journal* 120, no. 548: F452–F477.
Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora. 2013. Do labor market policies have displacement effects? Evidence from a clustered randomized experiment. *Quarterly Journal of Economics* 128, no. 2:531–80.
Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2009. glmnet: Lasso and elastic-net regularized generalized linear models. R package version.
Gautier, Pieter, Paul Muller, Bas van der Klaauw, Michael Rosholm, and Michael Svarer. 2012. Estimating equilibrium effects of job search assistance. IZA Discussion Paper 6748 (July), Institute for the Study of Labor, Bonn, Germany.
Gibbons, Robert, and Lawrence F. Katz. 1991. Layoffs and lemons. *Journal of Labor Economics* 9, no. 4:351–80.
Hoffman, Mitch, Lisa B. Kahn, and Danielle Li. 2015. Discretion in hiring. NBER Working Paper no. 21709, National Bureau of Economic Research, Cambridge, MA.
Horton, John. 2015. Price floors and preferences: Evidence from a minimum wage experiment. Working paper, New York University.
Horton, John, and Ramesh Johari. 2015. At what quality and what price? Eliciting buyer preferences as a market design problem. Paper presented at the 16th ACM Conference on Economics and Computation, Portland, OR.
Kluve, Jochen. 2010. The effectiveness of European active labor market programs. *Labour Economics* 17, no. 6:904–18.
Kuhn, Peter, and Hani Mansour. 2014. Is Internet job search still ineffective? *Economic Journal* 124, no. 581:1213–33.
Kuhn, Peter, and Mikal Skuterud. 2004. Internet job search and unemployment durations. *American Economic Review* 94, no. 1:218–32.
Oyer, Paul, and Scott Schaefer. 2011. Personnel economics: Hiring and incentives. In *Handbook of labor economics*, vol. 4b, ed. Orley Ashenfelter and David Card. Amsterdam: Elsevier.
Pallais, Amanda. 2014. Inefficient hiring in entry-level labor markets. *American Economic Review* 104, no. 11:3565–99.
Pellizzari, Michele. 2011. Employers' search and the efficiency of matching. *British Journal of Industrial Relations* 49, no. 1:25–53.
Phillips, David C. 2015. Search externalities and bias in correspondence experiments: Evidence from randomly assigned applicant pools. Working paper, Columbia University.

Resnick, Paul, and Hal R. Varian. 1997. Recommender systems. *Communications of the ACM* 40, no. 3:56–58.

Stanton, Christopher, and Catherine Thomas. 2012. Landing the first job: The value of intermediaries in online hiring. SSRN 1862109.

Tibshirani, Robert. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58, no. 1: 267–88.

van Ours, Jan, and Geert Ridder. 1992. Vacancies and the recruitment of new employees. *Journal of Labor Economics* 10, no. 2:138–55.

Varian, Hal R. 2010. Computer mediated transactions. *American Economic Review* 100, no. 2:1–10.