DISCUSSION PAPER SERIES

IZA

IZA DP No. 9501

Can Reputation Discipline the Gig Economy? Experimental Evidence from an Online Labor Market

Alan Benson
Aaron Sojourner
Akhmed Umyarov

November 2015

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Can Reputation Discipline the Gig Economy? Experimental Evidence from an Online Labor Market

Alan Benson
University of Minnesota

Aaron Sojourner
University of Minnesota and IZA

Akhmed Umyarov
University of Minnesota

Discussion Paper No. 9501
November 2015

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 9501
November 2015

ABSTRACT

Can Reputation Discipline the Gig Economy? Experimental Evidence from an Online Labor Market*

In two experiments, we examine the effects of employer reputation in an online labor market (Amazon Mechanical Turk) in which employers may decline to pay workers while keeping their work product. First, in an audit study of employers by a blinded worker, we find that working only for good employers yields 40% higher wages. Second, in an experiment that varied reputation, we find that good-reputation employers attract work of the same quality but at twice the rate as bad-reputation employers. This is the first clean, field evidence on the value of employer reputation. It can serve as collateral against opportunism in the absence of contract enforcement.

JEL Classification: L14, M55, J41, J2, L86, D82, K12, K42

Keywords: labor, personnel, contracts, online labor markets, job search, screening, reputation, online ratings

Corresponding author:

Aaron J. Sojourner
Carlson School of Management
University of Minnesota
321 19th Ave S 3-300
Minneapolis, MN 55455
USA
E-mail: asojourn@umn.edu

* Authorship is equal and alphabetical. We thank our excellent research assistants: Harshil Chalal, Sima Sajjadiani, Jordan Skeates-Strommen, Rob Vellela, and Qianyun Xie. For their useful feedback, we thank John Budd, Eliza Forsythe, Mitch Hoffman, Colleen Manchester, Mike Powell, David Rahman, Chris Stanton, and workshop participants at the 2015 ASSA Meetings, 2015 joint MEA-SOLE meetings, University of Minnesota Applied Microeconomics Workshop, MIT Sloan Organizational Economics Lunch, and the MIT Conference on Digital Experimentation. We also thank the University of Minnesota Social Media and Business Analytics Collaborative for funding.

Uber, Airbnb, TaskRabbit, and other online platforms have drastically reduced the price of microcontracting and prompted the birth of a "gig" economy. This transition has posed dilemmas for regulators as legacy operators allege these services circumvent regulations that protect service providers and consumers. Online platforms respond that their bilateral ratings systems discipline trading partners who break rules and norms, making traditional licensing and enforcement unnecessary. This dilemma is also playing out in online labor markets, where oDesk and eLance (now Upwork) developed monitoring and rating systems to discipline trading parties. In contrast, Amazon Mechanical Turk (M-Turk) features neither; after workers put forth effort, employers may keep the work product but refuse payment for any reason or no reason. Workers have no contractual recourse. A U.S. Government Accountability Office (2015, p. 22) report notes that such "online clearinghouses for obtaining ad hoc jobs" are attempting "to obscure or eliminate the link between the worker and the business..., which can lead to violations of worker protection laws."

Incomplete contracting, weak access to enforcement, and the disciplining role of reputation are not new to labor markets. A large literature considers the employers problem of identifying good workers, and their use of credentialing institutions like higher education to screen workers. While economists and legal scholars have long considered the reverse problem and theorized that reputational concerns constrain employer opportunism, there remains scant empirical work (Oyer and Schaefer, 2011). For workers, this is a dilemma because two prospective employers that offer identical employment contracts may actually differ widely in the criteria they apply for raises, promotions, terminations, scheduling, bonuses, task assignment, and many other working and payment conditions. In contingent, undocumented, and low-wage labor markets, concerns are as basic as whether employers will pay for all hours worked or pay at all. That jobseekers care about employers' reputations, or that employment contracts are incomplete, is self-evident.[1]

[1]Employers advertise favorable, "Best Place to Work" rankings in their recruitment materials. Jobseekers lean on experienced employees, professional associations, labor unions, word of mouth, and other signals to get a better understanding of employers' promotion and termination criteria, training opportunities, bonuses, flexibility, respectfulness, and other uses of discretionary authority. Several websites now enable workers to share experience with their employers, including Glassdoor, Careerbliss, Contratados, RateMyEmployer, eBossWatch, JobAdviser, Kununu, JobeeHive, TheJobCrowd, Ratemycompany, and the Freelancers Union's Client Scorecard. When a jobseeker asks, "How is your company to work for?" or a friend asks, "How's your new boss?" it would be obtuse to answer, "Here, read my employment contract." These questions attempt to uncover difficult-to-enforce aspects of the employment relationship. Models that assume workers have perfect information about employer heterogeneity gloss over the difficulty workers face in navigating these matters.

However, the ability to use online labor markets as a setting for clean experimental research on the value of an employer's reputation in labor markets is new, and for 40% of workers that the U.S. Government Accountability Office (2015) reports are now employed in contingent work arrangements, it is also potentially important. Indeed, the foundational empirical study on reputation and collective retribution in labor markets offered by (Greif, 1993) looks back to the 11th-century Maghrib as a setting where a coalition arose to discipline trading partners where contracts were not enforceable, though the accuracy of the historical account has been called into question Edwards, and Ogilvie (2008). In contrast, online labor markets present a contemporary and growing population, one that features well-defined inaccess to enforcement, and one that demands the question: can workers aggregate their private experiences into shared memory in order to discipline opportunistic employers?

When an employer values its credibility among its own workers, self-enforcing relational contracts may deter such opportunistic behavior. For example, if employer reneges on noncontractible subjective bonuses, its employees may discount the promise of future effort-continent bonuses (Baker, Gibbons and Murphy, 2002; Brown, Falk and Fehr, 2004). However, this mechanism focuses on incumbent workers accruing private information about their employer through personal experience and deciding whether to leave. For jobseekers lacking experience with an employer, can workers aggregate their private experiences into shared memory in order to discipline opportunistic employers?

M-Turk specifically has many features making it attractive for studying how workers navigate employer heterogeneity. First, there is no variation in the extent of contracts. In M-Turk, after workers put forth effort, employers may keep the work product but refuse payment for any reason or no reason. Workers have no contractual recourse. This complete lack of contract enforcement is rare and valuable. In most labor markets, relationships embody a mix of enforceable and unenforceable elements and the nature of the mix is unknown to the econometrician. Observed differences between employers may reflect differences in workers' access to legal recourse. Here, we know all employer behavior reflects discretionary action absent the possibility of enforcement. Second, M-Turk does not have a native employer-reputation system, a feature it shares with offline labor markets but unlike other online labor markets. This also proves useful by allowing us to uncouple worker effort from employer reputation in a part of the study.

To avoid employer opportunism, many M-Turk workers use Turkopticon, a third-party browser plugin that allows workers to review and screen employers. There are several reasons these ratings may be uninformative. First, the system is unnecessary if workers face no information or enforcement problem. Second, the system relies on workers voluntarily contributing accurate, private information to a common pool, which costs time and directs other workers to scarce, high-paying tasks. This distinguishes labor markets from consumer markets where trade is non-rival. Third, ratings systems vary widely in their informativeness due to reputation inflation and other issues (Nosko and Tadelis, 2014; Horton and Golden, 2015). Anyone can post any review on Turkopticon. It has no revenue and is maintained by volunteers.

In two experiments, we show that (1) employer reputations have value for workers, who use it to screen employers on otherwise-unobservable heterogeneity, and (2) to employers who can benefit when a better reputation makes it easier to attract more workers of any given quality, basically shifting out the labor supply curve they face. To our knowledge, these experiments provide the first estimates of the value of employer reputation measured in the field based on any design more credible than a control function.

The first experiment tests the validity of the online reputations from the perspective of a worker. We act as a worker to assess the extent to which other workers' public ratings reflect real variation in employer and job quality. One research assistant (RA) randomly selects tasks from employers who have good reputations, bad reputations, or no reputation and sends them to a second RA who is blind to employers' reputations. We find that effective wages while working for good-reputation employers is 40 percent greater than effective wages while working for bad-reputation employers.

The second experiment measures the effect of employers' reputations on their ability to recruit workers. We create 36 employers on M-Turk. Using Turkopticon, we endow them with (i) 8-12 good ratings, (ii) 8-12 bad ratings, or (iii) no ratings. We then examine the rate they attract workers to posted jobs. We find that employers with good reputations attract work about 50 percent more quickly than our otherwise-identical employers with no ratings and 100 percent more quickly than those with bad reputations. Using estimates of M-Turk wage elasticities published elsewhere, we estimate that posted wages would need to be almost 200 percent greater for bad-reputation employers and 100 percent greater for no-reputation employers to attract workers at the same rate as good-reputation employers do. Outside of M-Turk, one might think of the attractiveness of the job as the firm's ability to attract applicants and reputation as a substitute for wage for that purpose. We also estimate that about 55 percent of job-searchers use Turkopticon, suggesting that more complete adoption would magnify effects. We find evidence that Turkopticon is signaling employer characteristics rather than just task characteristics. These results demonstrate that workers use reputations to screen employers and that reputation affects employers' abilities to attract workers.

We propose a simple, equilibrium-search model consistent with our results. In the model, informed-type workers screen employers with bad reputations, and the threat of losing a good reputation and thus losing the informed workers discourages employers from engaging in wage theft and other forms of opportunism. The model depends crucially on the willingness of workers to provide accurate ratings that reflect employers' behaviors. In this way, employers' worker-created reputation serves as collateral against wage theft, effectively substituting for the role that formal contracts normally play in the labor market.

Turkopticon and other sites that diffuse workers' private information demonstrate the willingness of anonymous workers from diverse backgrounds to contribute to the collective punishment of employers who abuse an absence of contractual enforcement. As such, the two experiments and the model illustrate the value of an employer-reputation system for the workers who rely on it to identify good employers, for the good employers who rely on it to attract workers, and for the whole market which relies on it to solve the hold-up problem.

I M-Turk and Employer Reputation

M-Turk is an online labor market that allows employers (known as requesters) to crowdsource human intelligence tasks (HITs) to workers over a web browser. Common HITs include audio transcription, image recognition, text categorization, and other tasks not easily performed by machines. Amazon does not generally publish detailed usage statistics; however, in 2010, it reported that more than 500,000 workers from over 190 countries were registered on M-Turk.[2] In 2014, Panos Ipeirotis's web crawler found that the number of available HITs fluctuated between 200,000 and 800,000 from January and June 2014.[3] Ross et al. (2009) found that a majority of workers were female (55%) and from the U.S. (57%) or India (32%). Horton and Chilton (2010) estimates that the median reservation wage was

[2]Available online at https://forums.aws.amazon.com/thread.jspa?threadID=58891
[3]Available online at http://mturk-tracker.com (accessed June 14, 2014).

$1.38 an hour. M-Turk's revenue comes from 10% brokerage fees paid for by employers.

When an employer posts a task, it appears to workers on a list of available tasks. This list specifies a short description of the task, the number of tasks available in the batch, the promised pay per task, the time allotted for workers to complete the task once they accept it, and the name of the employer. The employer also may restrict eligibility to workers with a sufficiently high approval rating, which requires a history of having submitted work approved and paid for by past employers. Workers may preview the task before accepting. Upon acceptance, a worker has the allotted time to submit the task. The employer then has a predetermined period to approve or reject the task, with or without an accompanying note. If the employer approves the task, the employer pays the posted rate and broker fees to Amazon. The conditions for approval are not contractible; if the employer rejects the task, the worker's submitted work remains in the employer's possession but no payment is made. Moreover, the worker's approval rate will decline, reducing the worker's eligibility for other tasks. There is no process for appealing a rejection.

Opportunism takes many forms in this market. Employers may disguise wage theft by posting unpaid trial tasks, implicitly with the promise that workers who submit work that matches a known, correct answer will receive work for pay, when in fact the trial task is the task itself and the employer rejects all submitted work for being defective. In addition to nonpayment, employers may also advertise that a task should take a set amount of time when it is likely to take much longer. Therefore, although the promised pay for accepted submissions is known, the effective wage rate, depending on the time it takes to complete the task, is not. Employers can also delay accepting submitted work for up to thirty days. Employers may or may not communicate with workers.

Within M-Turk, there is no tool allowing workers to review employers, and workers cannot observe employers' effective wages or payment histories. However, several online, third-party resources have sprung up that allow workers to share information voluntarily regarding employer quality. These include web forums, automatic notification resources, and public-rating sites.[4]

This paper experimentally studies the value of reputation on Turkopticon, a community ratings database and web-browser plugin.[5] The plugin adds information to the worker's job

[4]Popular resources include CloudMeBaby.com, mturkforum.com, mturkgrind.com, turkalert.com, turkernation.com, turkopticon.ucsd.edu, and Reddit's HitsWorthTurkingFor.

[5]For background on Turkopticon, see (Silberman et al., 2010; Irani, 2012; Silberman, 2013).

search interface, including community ratings of an employer's communicativity, generosity, fairness, and promptness. Ratings take integer values from one to five. As of November 2013, Turkopticon included 105,909 reviews by 8,734 workers of 23,031 employers. The attributes have a mean of 3.80 and a standard deviation of 1.72.[6] Workers can click on a link to read text reviews of an employer. These reviews typically further recommend or warn against doing work for a given employer. Figure 1 provides an illustration.

[FIGURE 1]

[FIGURE 2]

Figures 1 and 2 illustrate an M-Turk worker's job search process. Figure 1 shows how workers search for tasks for pay. Figure 2 shows a preview of the task that we use for this study.

The information problem in this setting is related to the relational contracting literature (Baker, Gibbons and Murphy, 2002; Bull, 1987; Klein and Leffler, 1981; Telser, 1980). In the classic model, workers and firms accurately observe each other's past behavior and choose whether to cooperate beyond contractual obligations; the threat of future noncooperation sustains efficient cooperation. However, public reputation systems can facilitate the diffusion of (mis)information in the context of job search where firms and workers lack prior personal, bilateral experience.

Turkopticon is remarkable because it relies on voluntary feedback from a community of anonymous workers to provide a signal of employer quality. These reviews are costly in terms of the worker's time and the content of the review is unverifiable to other workers. More importantly, there is wide variation in the effective pay rate of individual tasks. Because employers typically post tasks in finite batches and allow workers to repeat tasks until the batch is completed, the wage-maximizing behavior would be to hoard tasks posted by

[6]These statistics are based on our analysis of data scraped from the site. Attribute ratings are determined by the mean from the following questions: (i) for communicativity, "how responsive has this requester been to communications or concerns you have raised?" (ii) for generosity, "how well has this requester paid for the amount of time their HITs take?" (iii) for fairness, "how fair has this requester been in approving or rejecting your work?" (iv) for promptness, "how promptly has this requester approved your work and paid?" Their means (standard deviations) are respectively 4.01 (1.68), 3.98 (1.62), 3.71 (1.68), and 3.18 (1.91), suggesting that ratings are meaningfully spread. Their number of reviews are 93,596, 93,025, 99,437, and 44,298. Reviews are somewhat consistent across dimensions; the correlation between any one dimension and the mean value of the other three dimensions is 0.57. On workers' displays, average ratings are color coded; scores less than 2 are red, scores between 2 and 3 are yellow, and scores greater than 3 are green.

good employers by misdirecting other workers.[7] Because reviews are anonymous, direct reciprocity and punishment is limited. As such, sharing honest reviews could be thought of as a prosocial behavior that is costly to the worker in terms of time and valuable private information, and in which social recognition or direct reciprocity is limited. Other studies of online reputation systems suggest that reviewers are primarily motivated by a "joy of giving" and fairness (Cornes and Sandler, 1994; Resnick and Zeckhauser, 2002).

Much of the theoretical work on reputation has focused on the reputation of sellers of goods, rather than employers as the purchasers of labor. Following Klein and Leffler (1981), theoretical work proposes that sellers with good reputations will be able to charge higher prices. In their study of eBay sellers, Bajari and Hortaçsu (2003) find only a small effect of reputation on prices. However, Banerjee and Duflo (2000) find that supplier reputation is important in the Indian software market, where postsupply service is important but difficult to contract. McDevitt (2011) finds evidence that residential plumbing firms with high records of complaints are more likely to change their name, suggesting that firms seek to purge bad reputations. MacLeod (2007) concludes that the evidence that reputation substitutes for prices is mixed.

M-Turk workers are unconventional, in that they're contracted for very small tasks and have minimal interaction with firms. However, the issues that they confront are more general. Where contractual protections are slim, wage theft substantially impacts earnings especially among independent contractors, undocumented immigrants, misclassified employees, and low-wage employees (Bobo, 2011; Rodgers, Horowitz and Wuolo, 2014). "Wage theft" has prompted the United States Wage & Hour Division to award back pay to an average of 262,996 workers a year for the past ten years, and far more cases go unremedied (Bernhardt, Spiller and Theodore, 2013; Bobo, 2011; Lifsher, 2014; Bernhardt et al., 2009; United States Department of Labor Wage and Hour Division, 2014). Where employment contracts are enforceable, they are rarely complete; terms such as degree of training, task assignments, promotion criteria, and termination criteria are difficult to contract.

However, empirical research on employer reputation as a deterrent to opportunism is slim. In a series of laboratory studies, Bartling, Fehr and Schmidt (2012) find that test subjects posing as employers are less likely to hold up those posing as workers when the experimenter will make their past actions observable to those same workers in future periods.

[7] This competition between workers to get the best jobs is the basis of resources such as TurkAlert.com, which allows workers to receive an alert whenever employers of their choosing post new tasks.

As predicted by relational contracting theory, private bilateral reputations develop and the prospect of lost value can deter employers from abusing authority. In their conclusion, they point to the potential value of a *public* reputation system, "it may be possible to improve the principals' incentives to acquire a good reputation by, for example, creating an institution that provides public information about the principals' reputation," though this lies outside the scope of their study.

While other studies have sought to identify the value of employer reputation outside the lab, identifying credibly-exogenous variation in employers' reputations has proven difficult. Turban and Cable (2003) provided the first correlational evidence that companies with better reputations tend to attract more applicants using career-services data from two business schools. Brown and Matsa (2013) find that distressed financial firms attract fewer and lower quality applicants. Hannon and Milkovich (1995) find mixed evidence that news of prominent employer rankings affects stock prices. Using a similar methodology, Chauvin and Guthrie (1994) find small but significant effects. While these two studies test the business value of good employer reputations, and they do so using institutions that arose organically, these specific methodologies are challenging to implement due to relatively low signal-to-noise ratios and small sample sizes. In these and the lab studies, reputation consists of some third-party signal rather than public, voluntary cheap talk among workers who share their private experiences.

Prior work in online labor markets has focused on the employers' problem of screening workers, rather than vice versa. Consistent with employer learning models, Pallais (2015) shows that prior work experience greatly improves workers' prospects for receiving job offers and higher pay. Agrawal, Lacetera and Lyons (2013) find that such experience is particularly beneficial for applicants from less developed countries, particularly among experienced employers. Stanton and Thomas (2014) find that outsourcing agencies help novice online workers signal their ability.

II Experiment 1

The first experiment examines the value of the reputation system to workers. Specifically, we examine whether Turkopticon ratings are informative of three employer characteristics that workers value but about which they face uncertainty during the search process: the likelihood of payment, the time to payment, and the implicit wage rate. As reflected in the literature on online ratings, informedness shouldn't be taken for granted. Horton and Golden (2015) show that oDesk, an online labor market with a native bilateral rating system, experiences extensive reputation inflation as employers and workers strategically, rather than truthfully, report experiences. Others report similar biases on eBay (Dellarocas, and Wood, 2008; Nosko and Tadelis, 2014), Airbnb (Fradkin et al., 2014), and Yelp (Luca and Zervas, 2014). The validity of Turkopticon ratings may be even more surprising, given that tasks offered by revealed good employers are rival (unlike, for example, good products on retail markets).

We follow the following procedure:

1. We produce a random ordering of three reputation types: Good, Bad, and None.

2. The nonblind research assistant (RA1), using a browser equipped with Turkopticon, screens the list of tasks on M-Turk until finding one that meets the requirements of the next task on the random ordering.

    - If the next scheduled item is Good, RA1 searches the list for a task posted by an employer in which all attributes are green (all attributes are greater than 3.0/5). 26.3% of the 23,031 employers reviewed on Turkopticon meet this criterion.
    - If the next scheduled item is Bad, RA1 searches the list for a task posted by an employer with no green attributes and a red rating for pay (all attributes are less than 3.0/5, and pay is less than 2.0/5). 21.6% of employers reviewed on Turkopticon meet this criterion.
    - If the next scheduled item is None, RA1 searches the list for a task posted by an employer with no reviews.

3. RA1 sends the task to the blinded RA2, who uses a browser not equipped with Turkopticon.

4. RA2 performs and submits the task. RA2 is instructed to perform all tasks diligently.[8]

[8] RA2 was not able to complete all jobs sent by RA1. Some expired quickly. Also, bad-reputation employers' jobs were more likely to be so dysfunctional as to be unsubmittable.

5. RA1 and RA2 repeat steps 2-4. A web crawler records payments and rejections by employers to RA2's account with accuracy within 1 minute of actual payment or rejection.

The blinding procedure decouples the search process from the job performance process, thereby protecting against the risk that RA2 inadvertently conditions effort on the employer's reputation.

[FIGURE 3]

Figure 3 shows results for rejection rates and time-to-payment by the employer's reputation type. Rejection rates were 1.4 percent for employers with good reputations, 4.3 percent for employers with no reputation, and 7.5 percent for employers with bad reputations.

[TABLE 1]

Table 1 presents further results and significance tests for rejection rates, time-to-payment, and realized hourly wage rates. We define realized wage rates to be payments divided by the time to complete the task if the work is accepted and zero if the work is rejected. We define promised wage rates to be posted payments divided by the time to complete the task; they are not zero if the work is rejected.[9] Employers with good reputations have significantly lower rejection rates and faster times-to-decisions. They do not have statistically different posted pay rates. This distinction is important because the pay for accepted tasks is contractible but the task's acceptance criteria and realistic time requirements are not.

In principle, the ratings on Turkopticon could be orthogonal to employer type, and instead be providing information on task types (e.g. survey or photo categorization) rather than employer types. We do not find evidence that this is the case. First, Turkopticon requests workers to rate employers on fairness, communicativity, promptness, and generosity; unlike task type, these are revealed only after workers have invested effort and are subject to hold-up. Textual comments also emphasize information that would only be revealed to prospective workers after investing effort. Second, the RA's task classifications in experiment

[9]Counts are lower for wage rates because the blinded RA lost track of time-to-completion for some tasks.

1 are not significantly correlated with Turkopticon scores. We also test for evidence of task screening in experiment 2.

Given the low cost of creating new employers, it is puzzling that employers with poor reputations persist. When the study was conducted, the only cost to creating a new employer was the time filling forms and awaiting approval. Since then, the cost of producing new aliases has grown.[10] If creating new accounts were perfectly costless and employers were informed, we would expect there to be no active employers with poor reputations. However, Turkopticon's textual reviews also suggest that workers are aware that employers with bad reputations may create new identities.

We conclude that the longer work times and lower acceptance rates validate the ratings as informative about employer differences that would be unobservable in the absence of the reputation system.

To provide an intuition for the magnitude of the value of employer-reputation information to workers, note that our results imply that following a strategy of doing jobs only for good-reputation employers would yield about a 40 percent higher effective wage than doing jobs only no-reputation or bad-reputation employers: \$2.83 versus just under \$2.00 per hour. Results suggest about 20% of the gap in effective pay is explained by nonpayment and 80% is explained by longer tasks. However, this calculation understates the penalties when an employer rejects tasks because the rejected worker is penalized in two ways: nonpayment and a lower approval rating. The latter reduces the worker's eligibility for future tasks from other employers.

III Experiment 2

The second experiment examines the value of the reputation system to employers. Specifically, we examine whether a good reputation helps employers attract workers. We do so by creating employers on M-Turk, exogenously endowing them with reputations on Turkopticon, and then testing the rate at which they attract work.

1. We create 36 employer accounts on M-Turk. The names of these employers consist of permutations of three first names and twelve last names.[11] We use multiple employers

[10] On July 27, 2014, Amazon began requiring employers to post a legal personal or company name, physical address, and Social Security Number or Employer Identification Number.

[11] The first names are Joseph, Mark, and Thomas. The last names are Adams, Clark, Johnson, Jordan,

to protect against the evolution of ratings during the experiment. We choose these names because they are: common, Anglo, male (for first names), and our analysis of Turkopticon ratings find that these names are not generally rated high or low.

2. We endow 12 employers with good reputations and 12 employers with bad reputations. We do so by creating accounts on Turkopticon and posting numerical attribute ratings and longform text reviews. Reviews for our bad-(good-)reputation employers are taken as a sample of actual bad(good) reviews of bad-(good-)reputation employers on Turkopticon.[12] Good- and bad-reputation employers receive eight to twelve reviews each. Because M-Turk workers may sort tasks alphabetically by requesters' names, we balance reputations by the first name of the employer so that reputation is random with respect to the alphabetical order of the employer.

3. Our employer identities take turns posting tasks on M-Turk. They do so in seventy-two one-hour intervals, posting new tasks on the hour. Posts began at 12:00 AM on Tuesday, July 7 and ended at 11:59 PM on Thursday, July 9. For example, the employer named Mark Kelly, who was endowed with a good reputation on Turkopticon, posted tasks at 12:00 AM and ceased accepting new submissions at 12:59 AM, thereafter disappearing from workers' search results. At 1:00 AM, Joseph Warren, who had no reputation on Turkopticon, posted new tasks.

   We balance the intervals so that: (1) in each hour, over three days, the three reputation types are represented once, (2) in each hour, over each six-hour partition of a day, the three reputation types are represented twice. We chose the final schedule (Appendix Table 5) at random from the set of all schedules that would satisfy these criteria.

   The tasks consist of image recognition exercises. Workers are asked to enter the names, quantity, and prices of alcoholic items from an image of a grocery receipt that we generated. Receipts are twenty items long and contain three to five alcoholic items.[13] Workers may only submit one task in any one-hour interval. The pay rate is $0.20, and workers have fifteen minutes to complete the task once they accept it.

Kelly, Lewis, Martin, Miller, Owens, Roberts, Robinson, and Warren.

[12]For this purpose, we define bad reviews as those giving a score of 1/5 on all rated attributes and a good review as giving a 4/5 or 5/5 on all rated attributes. The text reviews clearly corroborate the numerical rankings; an RA given only the text reviews correctly identified the employer type in 285 of the 288 reviews.

[13]Alcoholic items came from a list of 25 bestselling beers. This task therefore features simple image recognition, abbreviation recognition, and domain knowledge.

4. Simultaneously, we create three employers that post 12-cent surveys requesting information from workers' dashboards. These employers post new batches of tasks each hour for twenty-four hours each. Their reputation does not vary. The purpose of this task is to determine a natural baseline arrival rate that could be used as a control in the main regressions.

5. We record the quantity and quality of completed tasks. We do not respond to communications and do not pay workers until the experiment concludes.

As a study of employer reputation, we anticipated that reputation may evolve naturally over the course of the experiment as workers discussed the tasks on public forums. If reputation propagated from Turkopticon to other forums, we expected the effect of reputation to rise over time. If workers noticed and publicized that employers of different names actually had the same identity, we expected the result to diminish over time.

The first instance occurred at 7 PM on Tuesday, when a task was recommended on the Reddit subforum "HITs Worth Turking For."[14] On Thursday[15] at 4:14 PM, a worker posted a list of the 24 employers with good and bad ratings on Reddit, noting their similarities and suggesting that the reviews were created by fake accounts. On Thursday at 5:22 PM, to address concerns that employers were falsifying reviews with the intent of defrauding workers, we announced the experiment to a concerned group of workers on a Turkopticon discussion board and disclosed that all workers would be paid. On Thursday at 6:14 PM, the description of the experiment was cross-posted on Reddit.

Summarizing the results of the experiment, Figure 4 shows the cumulative distribution of arrivals across the three employer reputation types. By the conclusion of each of the twelve six-hour partitions, the employer with good ratings had attracted more work than the employer with neutral ratings, and the employer with neutral ratings had attracted more work than the employer with poor ratings.

[FIGURE 4]

Table 2 shows results from a Poisson regression model. Poisson regression results find that the differences in the arrival rates of submitted tasks are generally statistically

[14]The post included a link to the task and the note: "Similar to the ones posted earlier, entering alcoholic purchases from a receipt. Takes less than a minute, excellent [Turkopticon rating]."

[15]Thursday is the last day of the three days of the experiment

significant across partitions of the experiment. They are also robust to day and hour fixed effects, and to using the baseline task's arrival rate as a control. The arrival rate of task previews, task acceptances, and error-free submissions was also significantly faster for the employer with a good reputation and slower for the employer with a poor reputation.

[TABLE 2]

Table 3 shows results from a negative binomial model. This allows for overdispersion, relaxing the Poisson regression assumption that counts follow a Poisson distribution with $E(Y) = Var(Y)$.[16] These regressions generally reject that counts follow a Poisson distribution, leading us to prefer the negative binomial model.

In all samples except for the six-hour partitions, employers with good reputations attract work more quickly than employers with poor reputations with $p < 0.01$. However, if comparing only against no-reputation employers at a 5% significance level, employers with a good reputation do not receive submitted work significantly faster than those with no reputation, and employers with a poor reputation receive submitted work significantly slower only in the full samples.

[TABLE 3]

We also examine differences in estimated effort and quality. The mean time spent per task for good reputation, no reputation, and poor reputation employers were respectively 136, 113, and 121 seconds. The difference between good reputation and no reputation employers is statistically significant with $p < 0.01$. For each of the three groups, the error-free rates were between 61% and 63% and the major-error rates (e.g. no alcoholic items identified) were between 3.0% and 5.2%. Differences in the error-free rates and major-error rates are not statistically significant.[17] Mason and Watts (2009) also found that higher payments raise the quantity, but not quality, of submitted work; it appears to be difficult to improve quality by either reputation or pay.[18]

[16] Overdispersion may have resulted from time-of-day effects.

[17] Differences are for a two-sample t-test for equal means of the log-work time with $\alpha < 0.1$. Error-free receipts are those in which all alcoholic items were identified, no non-alcoholic items were identified, and the prices were entered correctly. Major-error receipts are those in which no alcoholic items were identified, or more than six items are listed.

[18] However, by attracting the same amount of work at a lower pay, good reputation employers may presumably purchase higher quality by duplicating tasks and adopting a "majority rules" policy. As such, quantity and quality at any given pay level may be thought of as substitutes, and results suggest that

In the full sample, 45.2% of the submitted tasks were not the first tasks submitted by an individual worker, and 9.7% of the submitted tasks were the sixth task or greater. The high incidence of repeat submissions may be for a number of factors, including: power-users, correlated task search criteria (e.g. individuals continuously search using the same criteria), automated alerts (e.g. TurkAlert), or purposely searching for the same task across hours.

Table 4 shows results from our preferred specification of the negative binomial regressions to estimate the arrival rates of task previews, acceptances, submissions, first submissions (by worker), and correct first submissions. These specifications omit the last twelve hours in which the experiment was disclosed and also include day and hour fixed effects. Arrival rates for good reputation employers are significantly greater than no reputation employers for all outcomes, and arrival rates for no reputation employers are significantly greater than bad reputation employers for all outcomes except correct first submissions with $p < 0.05$. Results provide evidence that good reputations produce more previews, acceptances, submissions, first submissions, and correct first submissions.

[TABLE 4]

The point estimates in column (3) suggest employers with good and no reputations respectively outperform those with bad reputations by 84% and 36%. Horton and Chilton (2010) estimate that M-Turk workers have an extensive-margin, median-wage elasticity of 0.43. If this point elasticity holds for our sample, a bad-reputation employer that pays \$0.59, a no-reputation employer that pays \$0.37, and a good-reputation employer that pays \$0.20 would attract work at the same rate.

Table 4 also provides evidence about the effects of reputation on various steps in the matching process. Conditional on a worker previewing a task, the probability of accepting the task is not significantly different by treatment. If information received by previewing a task (e.g. the type of the task, the intuitiveness of the user interface) were a substitute for reputation information, then good reputation employers would lose fewer workers during the preview stage than no-reputation employers. In the former, but not latter, workers would already have received the signal prior to previewing the task. This evidence suggests that observable task characteristics do not substitute for reputation information. The reputation

employers with a good reputation may extract a higher quantity, *ceteris paribus*. Likewise, in the broader labor market, a good reputation may allow an employer to attract better applicants at any wage offer.

system adds information above what workers can otherwise observe.

Turkopticon is not native to the M-Turk interface and must be installed by the worker. As such, the reputations we endow are visible only to a fraction of workers, and so only part of the "treated" population actually receives the treatment. To estimate the share of M-Turk jobseekers who use Turkopticon, we posted a one-question, free response survey asking, "How do you choose whether or not to accept HITs from a requester you haven't worked for before? Please describe any factors you consider, any steps you take, and any tools or resources you use." Because we posted the survey from a requester account that did not have a Turkopticon rating, and because we require workers to identify Turkopticon specifically, we expected this procedure to yield a conservative estimate of the true portion of job-seekers who use Turkopticon. Of these, fifty-five of the 100 responses mention Turkopticon explicitly, and seven other responses mention other or unspecified websites.[19] To the extent the models estimate the effect of a known reputation on an employer's ability to attract work, we expect non-participation in Turkopticon to result in attenuation bias that would reduce the magnitude of coefficients and raise standard errors; adjusting for this attenuation bias would magnify estimates of the treatment effect by about 80% (0.55^{-1}). Naturally, this should be treated as a local prediction for the equilibrium we observe, and not a counterfactual rate for a scenario in which all workers use Turkopticon.

Experiment 2 also offers three additional pieces of evidence that Turkopticon provides information of employer type rather than task type. First, we find that observed probability of accepting a task conditional on previewing a task does not vary significantly by employer type. Second, we find that the elapsed time that workers spend previewing tasks prior to accepting the task does not vary significantly by reputation type. Third, our survey of 100 M-Turk workers featured no workers who reported a belief that certain tasks were inherently more fairly or highly compensated, though nearly all cited observable employer characteristics from past experience or tools like Turkopticon. These suggest that workers screened on Turkopticon ratings and not on information (e.g. task type) gathered during the task previews. This, along with ratings criteria used by Turkopticon and the test in Experiment 1, lead us to conclude that workers use Turkopticon to get information about employers that wouldn't be accessible until after they would have otherwise exerted effort (e.g. time to completion and nonpayment), rather than getting information on task type.

[19]Otherwise, responses emphasize estimated pay, estimated time to completion, and perceived trustworthiness (e.g. from a known organization).

Altogether, the second experiment supports the hypothesis that workers are attracted to employers with a good reputation and discouraged from those with a bad reputation. Through the experiment, the spread of information from Turkopticon to other sites also demonstrates how M-Turk workers use public forums to attract others to well-reputed employers.

IV Model

We offer a model of job search in which there is no contract enforcement and yet some employers are deterred from nonpayment by the threat of losing future work. Workers incur a search cost to receive a wage offer from a random employer. Some share of workers are "informed," able to observe any employer's pay history perfectly.[20] If the worker accepts the offer, the worker further incurs a cost of effort, produces work product, and then the employer chooses whether to pay or to renege. If the employer reneges, informed workers will refuse to work for them in the future. We take the share of informed workers to be exogenous, and characterize an interesting but non-unique equilibrium in which employers with a good reputation continue to pay as long as this share is sufficiently high. Otherwise, the reneging temptation is too great and all workers exit from the labor market.[21]

We refer to employers' practices of always paying or never paying as high-road and low-road strategies, and to the employers themselves as high-road and low-road employers. Low-road employers attract work more slowly but save on labor costs. High-road employers attract work more quickly but pay more in wages.[22] The share of low-road employers increases in the share of uninformed workers and the value created by a match. It decreases in the cost of search and the cost of worker effort.

[20]Perfect monitoring simplifies the exposition. Board and ter Vehn Moritz (n.d.) considers reputation building when learning is imperfect. Their model also yields ergodic shirking, with increasing incentives for noncontractible investments as reputation becomes noiseless.

[21]Other studies show how reputation systems and credentials can improve efficiency in other online markets including eBay (Nosko and Tadelis, 2014; Hui et al., 2014) and Airbnb (Fradkin et al., 2014).

[22]Workers may face the two standard kinds of information problems with respect to unobserved employer heterogeneity: adverse selection and moral hazard. Employers' technologies or product markets may differ in ways that make low-road practices more or less profitable. In this adverse-selection setting, it is trivial to understand why variation in employment practices emerges. An alternative theory is that there is no essential heterogeneity between employers. Differences in strategic employment practices appear between essentially-homogeneous employers. We focus on this, more-interesting case. In all labor markets, both mechanisms are almost certainly empirically relevant. Cabral and Hortaçsu (2010) did such an accounting in a consumer-goods market, baseball cards on EBay. We know of no analogous accounting in any labor market. That remains for future work.

Consider the following job search environment. There are measure 1 of workers indexed by $i \in [0, 1]$ and measure 1 of risk-neutral employers indexed by $j \in [0, 1]$. Workers with $i \leq p \in [0, 1)$ are informed to employers' past play. Workers who are indifferent between accepting and rejecting offers choose to accept. Employers indifferent between paying and reneging choose to pay. The timing of a period of job search follows:

1. Worker i chooses whether to search. Those who do incur cost c and receive a wage promise w from a random employer-j. Informed workers also observe j's past decisions to pay or renege. Non-searching workers receive 0 and proceed to the next period of job search. Think of 0 as the value of not participating in the labor market.

2. Worker i decides whether to accept or reject employer j's offer. If the worker accepts, he incurs cost of effort e and j receives work product with value y. If the worker rejects, he receives 0 and proceeds to the next period of job search.

3. Employer-j decides whether to pay w or to renege and pay 0. Employers discount future periods at rate δ.

To focus on the interesting equilibrium, suppose the following parameter restrictions. First, the gains from trade, farsightedness, and share of informed workers are sufficiently great that high-road employers do not renege, $\delta py - w \geq 0$. Second, promised wages and the share of high-road employers (denoted by $s \in [0, 1)$) are sufficiently great that workers participate in the labor market, $sw - c - e \geq 0$. Under these conditions, there exists an equilibrium in which:

1. For high-(low-)road employers it is incentive compatible in any period to (not) pay.

2. Informed workers employ a trigger strategy, accepting only offers from employers that have never reneged.

3. Uninformed workers accept all jobs.

4. The share of high-road employers will increase in the share of informed workers.

5. When the share of workers is nonzero, the arrival rate of work is greater for high road employers than for low-road employers.

Proof: Consider the case of a low-road employer. In any period, with probability p, the offer is received and rejected by an informed worker, yielding a payoff 0. With probability $1-p$, the offer is received and accepted by an uninformed worker, yielding payoff y. Low-road employers receive no benefit from paying wage w in any period. Then the arrival rate of accepted tasks for low-road employers is $1-p$ and the present value payoff is $(1-\delta)^{-1}(1-p)y$. Now consider high-road employers. In this case, all offers are accepted and all workers are paid, yielding an arrival rate of $1 \geq (1-p)$ and a present value payoff $(1-\delta)^{-1}(y-w)$. High-road employers prefer payment to reneging if $(1-\delta)^{-1}(y-w) \geq y + \delta(1-\delta)^{-1}(1-p)y$. Reducing yields the difference in present value of paying $\delta py - w \geq 0$, which follows from the first parameter restriction. Now consider workers. Informed workers encounter a high-road employer in any period with probability s. They accept offers from high-road employers because $w - e - c \geq -c$, which follows from $sw - c - e \geq 0$. They reject offers from low-road employers because $-c > -c - e$. Therefore, the present value of this strategy is $(1-\delta)^{-1}[s(w-e) - c]$. Uninformed workers accept all offers. Their present value is $(1-\delta)^{-1}(sw - e - c)$. Both informed and uninformed workers' payoffs satisfy their labor force participation constraint under the parameter restriction $sw - c - e > 0$.

The high-road employer's incentive compatibility constraint, $\delta py - w \geq 0$, is satisfied if three conditions are met: a sufficiently informed workforce would discipline a high-road employer that chose to renege, sufficiently farsighted employers that do not discount this punishment, and sufficient rents. Otherwise, high road employers choose instead to renege, the value of market participation for all workers becomes negative, and no work is performed.

The workers' participation constraint requires a sufficiently high share of employers that pay. Given p, the share of high-road employers (s) cannot fall below $\underline{s} \geq (e+c)(\delta py)^{-1}$. For low values of s, the payoff for uninformed workers does not satisfy their participation constraint. These conditions imply which combinations of worker-informedness p and high-road employer shares s are supportable in this equilibrium.

V Conclusion

Our main results provide evidence that reputation in M-Turk is valuable for both workers and for employers with good reputations. In our experiment, we get clean measures of the partial equilibrium values of employer reputation for workers and employers. Public, collectively-created reputation is valuable for workers because it lets them differentiate otherwise indistinguishable employers that in fact differ systematically.

We estimate that working only for good-reputation employers would make workers' wages about 40 percent higher than working for no- or bad-reputation employers. We find that good- and bad-reputation employers promise the same payments on average, consistent with a pooling equilibrium where bad-reputation employers try to blend in with good-reputation employers in the view of uninformed, jobseeking workers. However, bad-reputation employers' tasks take longer and they are far more likely to refuse to pay the worker.

Because many workers do use the reputation system in deciding whom to work for, employers with good reputations enjoy twice the arrival rate of bad-reputation employers. Average quality of work done by these newly-arrived workers does not differ by employer reputation. This should enable employers with better reputations to operate at a faster pace, a larger scale, or to be more selective in hiring.

M-Turk, like many microcontracting services, offers little contractual protection for workers. Payment for services, time to payment, and implicit wage rates are all noncontractible. However, this study demonstrates that workers contribute to a collective memory that serves to discipline and deter bad behavior. It also suggests that a well-managed reputation system may effectively substitute for such enforcement.

With its administrative data, Amazon could give workers access to historical information on each employer such as average past wage and rejection rates. It could also create a native, subjective rating system, as oDesk-Elance has and as Amazon has for consumer products. The lack of information about employer reputations coupled with the lack of contract enforcement may be limiting the market to the small size that a reputation can discipline, and to small tasks that are relatively short and well-defined; relatively few workers would risk investing a week into a task when the criteria for acceptance are poorly defined and payment is nonenforceable (Ipeirotis, 2010).

Some empirical results warrant future attention. First, why do workers rate employers? Because variation in realized wages is wide and tasks posted by good employers are scarce, revealed good employers could be thought of as valuable private information. Nevertheless, these ratings are informative. Workers may be motivated by altruism toward other workers, by altruism to good employers, or by a desire to punish bad employers. Second, in experiment 1, why did effective wages for good reputation employers exceed those for bad reputation employers? Following Klein and Leffler (1981), when there is a potential hold-up problem, good reputations should allow trading partners to extract favorable terms, such as the ability to attract work at lower pay. It's possible that an employer's reputation is correlated with other employer characteristics. One possibility, following Bartling, Fehr and Schmidt (2012), is that employers are heterogeneous in their altruism, and altruistic employers pay higher wages and have better reputations. Indeed, Turkopticon ratings include an item for generosity, which intends to capture expected wages. A second alternative is that employers are heterogeneous in their discount rates, and impatient employers pay higher wages and maintain good reputations to get work accomplished quickly. In M-Turk, these underlying employer characteristics may be more important than the mechanism offered by Klein and Leffler alone, and may also offer some guidance as to why Klein and Leffler's predictions have sometimes had mixed success empirically.

What relevance does this have for other labor markets? As on M-Turk, workers in the broader labor market strive to distinguish which employers will treat them well or ill. Workers have always made decisions with partial information about employer quality and, so, these forces have always shaped labor markets. Contracts and bilateral relational contracting are important forces disciplining employer opportunism, but they are undoubtedly incomplete. Workers have always relied on public employer reputations propagated through informal, decentralized, word-of-mouth conversations. Though economists have had theories about how employer reputation would work, the informal system has operated largely outside our view, yielding a very thin empirical literature. As the cost of communications and data-storage fell in recent years, employer reputation has showed up online in sites like Glassdoor. It has become more centralized, systematic and measurable. While this study develops the first clean evidence that an employer-reputation system affects labor-market outcomes, it will not be the last. New data on reputation in broader markets means other studies will follow. However, in the setting of conventional employment, it will be more challenging to shock reputation cleanly, to believe that unobservable channels of communication are not creating confounds, and to measure variation in outcomes such as wages, wage theft, and worker arrival rates.

Attention to the worker's information problem also suggests innovative directions for policy and institution-building. Can more be done to improve the functioning of the labor market through helping workers' overcome their information problem with respect to employer heterogeneity? Most markets have information problems to some degree. For M-Turk workers, Turkopticon is the Dun & Bradstreet of procurers, the Moody's of bond buyers, the Fair Isaac of consumer lenders, and the Metacritic of moviegoers. Each of these institutions offers extralegal protections to protect against contractual incompleteness based on information sharing and the implicit threat of coordinated withdrawal of trade by one side a market. A policy example of this kind of logic in action is that, the U.S. Occupational Safety and Health Administration began in 2009 systematically issuing press releases to notify the public about large violations of workplace safety laws. They attempt to influence employer reputation, to improve the flow of information about employer quality, and to create incentives for providing safer workplaces (Johnson, 2015). Workers have traditionally used labor unions and professional associations as a venue for exchanging information about working conditions and coordinating collective withdrawal of trade in order to discipline employers. The rise of new institutions that facilitate information sharing may be taking up some of this role.

References

Agrawal, AK, Nicola Lacetera, and Elizabeth Lyons. 2013. "Does Information Help or Hinder Job Applicants from Less Developed Countries in Online Markets?" National Bureau of Economic Research.

Bajari, Patrick, and Ali Hortaçsu. 2003. "The winner's curse, reserve prices, and endogenous entry: empirical insights from eBay auctions." *RAND Journal of Economics*, 34(2): 329–355.

Baker, George, Robert Gibbons, and KJ Murphy. 2002. "Relational Contracts and the Theory of the Firm." *Quarterly Journal of Economics*, 117(1): 39–84.

Banerjee, Abhijit V, and Esther Duflo. 2000. "Reputation effects and the limits of contracting: A study of the Indian software industry." *Quarterly Journal of Economics*, 115(3): 989–1017.

Bartling, Björn, Ernst Fehr, and Klaus M Schmidt. 2012. "Use and Abuse of Authority: A behavioral foundation of the employment relationship." IZA Discussion Paper #7029 4.

Bernhardt, Annette, et al. 2009. *Broken laws, unprotected workers: violations of employment and labor laws in America's cities.* A National Employment Law Project report. Available online at http://www.nelp.org/page/-/brokenlaws/BrokenLawsReport2009.pdf?nocdn=1.

Bernhardt, Annette, Michael Spiller, and Nik Theodore. 2013. "Employers gone rogue: Explaining industry variation in violations of workplace laws." *ILRReview*, 66(4): 808–832.

Board, Simon, and Meyer ter Vehn Moritz. n.d.. *Econometrica*, , (6).

Bobo, Kim. 2011. *Wage theft in America.* The New Press.

Brown, Martin, Armin Falk, and Ernst Fehr. 2004. "Relational contracts and the nature of market interactions." *Econometrica*, 72(3): 747–780.

Brown, Jennifer, and David Matsa. 2013. "Boarding a sinking ship? An investigation of job applications to distressed firms." Working paper. Available online at http://economics.mit.edu/files/8675.

Bull, Clive. 1987. "The existence of self-enforcing implicit contracts." *The Quarterly Journal of Economics*, 102(1): 147–160.

Cabral, Luis and Ali Hortaçsu 2010. "The Dynamics of Seller Reputation: Evidence from EBay." *Journal of Industrial Economics*, 58: 54–78.

Carmichael, HL. 1984. "Reputations in the labor market." *The American Economic Review*, 74(4): 713–725.

Chauvin, Keith W, and James P Guthrie. 1994. "Labor market reputation and the value of the firm." *Managerial and Decision Economics*, 15(6): 543–552.

Cornes, Richard, and Todd Sandler. 1994. “The comparative static properties of the impure public good model.” *Journal of Public Economics*, 54(3): 403–421.

Dellarocas, Chrysanthos, and Charles Wood. 2008. “The sound of silence in online feedback: Estimating trading risks in the prescence of reporting bias.” *Management Science*, 54(3): 460–76.

Edwards, Jeremy, and Sheilagh Ogilvie. 2008. “Contract Enforcement, Institutions and Social Capital: the Maghribi Traders Reappraised.” CES Working Paper No. 2254.

Estreicher, Samuel. 2009. “Employer Reputation at Work.” *Hofstra Labor & Employment Law Journal*, 27(1): 1–12.

Fradkin, Andrey, Elena Grewal, David Holtz, and Matthew Pearson. 2014. “Reporting Bias and Reciprocity in Online Reviews: Evidence From Field Experiments on Airbnb.”

Greif, Avner. 1993. “Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders’ Coalition.” *American Economic Review*, 83(3): 525-548.

Hannon, J, and GT Milkovich. 1995. “Does HR reputation affect shareholder value.” *Unpublished manuscript.*

Holmstrom, Bengt. 1981. “Contractual models of the labor market.” *The American Economic Review*, 71(2): 308–313.

Horton, John Joseph, and Lydia B Chilton. 2010. “The labor economics of paid crowdsourcing.” 209–218, ACM.

Horton, John, and Joseph Golden. 2015. “Reputation Inflation in an Online Market.” Working paper.

Hui, Xiang, Maryam Saeedi, Zeqian Shen, and Neel Sundaresan. 2014. “Reputation & Regulations: Evidence from eBay.”

Ipeirotis, Panos. 2010. “A Plea to Amazon: Fix Mechanical Turk!” *http://www.behind-the-enemy-lines.com/2010/10/plea-to-amazon-fix-mechanical-turk.html* Accessed: March 20, 2015.

Irani, Lilly. 2012. “Microworking the crowd.” *Limn*, 1(2).

Johnson, Matthew. 2015. “How the Cover Story Promotes Compliance.” *Boston College manuscript.*

Klein, Benjamin, and Keith B Leffler. 1981. “The role of market forces in assuring contractual performance.” *The Journal of Political Economy*, 89(4): 615–641.

Lifsher, Marc. 2014. “California cracks down on wage theft by employers.” *Los Angeles Times.*

Luca, Micael, and Georgios Zervas. 2014. “Fake it till you make it: Reputation, Competition, and Yelp Review Fraud.” *Harvard Business School NOM Unit Working Paper No. 14-006.*

MacLeod, W Bentley. 2007. "Can contract theory explain social preferences?" *The American Economic Review*, 97(2): 187–192.

Mason, Winter, and Duncan Watts. 2009. "Financial incentives and the 'performance of crowds.'" *SIGKDD Explorations*, 11(2): 100-108.

McDevitt, Ryan. 2011. "Names and reputations: an empirical analysis. *American Economic Journal: Microeconomics*, 3(3):193-209.

Nosko, Chris, and Steven Tadelis. 2014. "The Limits of Reputation in Platform Markets: An Empirical Analysis and Field Experiment."

Oyer, Paul, and Scott Schaefer. 2011. "Personnel economics: hiring and incentives." *Handbook of Labor Economics*, 4: 1769–1823.

Pallais, Amanda. 2015. "Inefficient hiring in entry-level labor markets." Forthcoming, *American Economic Review.*

Resnick, Paul, and Richard Zeckhauser. 2002. "Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system." *Advances in applied microeconomics*, 11: 127–157.

Ritter, Joseph and Lowell Taylor. 1994. "Workers as Creditors: Performance Bonds and Efficiency Wages." *American Economic Review*, 84(3): 694–704.

Rodgers, William M, Sara Horowitz, and Gabrielle Wuolo. 2014. "The Impact of Client Nonpayment on the Income of Contingent Workers: Evidence from the Freelancers Union Independent Worker Survey." *Industrial & Labor Relations Review*, 67(3 suppl): 702–733.

Ross, Joel, Andrew Zaldivar, Lilly Irani, and Bill Tomlinson. 2009. "Who are the turkers? worker demographics in amazon mechanical turk." *Department of Informatics, University of California, Irvine, USA, Tech. Rep.*

Silberman, M, Joel Ross, Lilly Irani, and Bill Tomlinson. 2010. "Sellers' problems in human computation markets." 18–21, ACM.

Silberman, M Six. 2013. "Dynamics and governance of crowd work markets." Workshop on Human Computation and Crowd Work.

Simon, Herbert A. 1951. "A formal theory of the employment relationship." *Econometrica: Journal of the Econometric Society*, 19(3): 293–305.

Stanton, Christopher, and Catherine Thomas. "Landing the First Job: The Value of Intermediaries in Online Hiring." Working paper.

Telser, L. G. 1980. "A Theory of Self-Enforcing Agreements." *The Journal of Business*, 53(1): 27.

Turban, Daniel B, and Daniel M Cable. 2003. "Firm reputation and applicant pool characteristics." *Journal of Organizational Behavior*, 24(6): 733–751.

United States Department of Labor Wage and Hour Division. 2014. "Fiscal Year Statistics for WHD." *http://www.dol.gov/whd/statistics/*, Accessed: Nov. 3, 2014.

U.S. Government Accountability Office. 2015. "Contingent Workforce: Size, Characteristics, Earnings, and Benefits." U.S. G.A.O. Report GAO-15-168R Continent Workforce.

Wachter, Michael L, and Randall D Wright. 1990. "The economics of internal labor markets." *Industrial Relations: A Journal of Economy and Society*, 29(2): 240–262.

VI Tables

Table 1— Rejection and time-to-payment by employer reputation

	Mean	Std. Error	N	paired test p-values Good	None	Bad
Main outcomes						
1. Rejection rates						
Good Reputation	0.013	0.008	223		0.073	0.003
No Reputation	0.043	0.016	164	0.073		0.246
Bad Reputation	0.071	0.018	211	0.003	0.246	
2. Days to decision						
Good Reputation	1.679	0.146	223		0.132	0.001
No Reputation	2.296	0.433	164	0.132		0.03
Bad Reputation	3.715	0.467	211	0.001	0.03	
3. Realized wage rates						
Good Reputation	2.834	0.228	173		0.011	0.043
No Reputation	1.957	0.259	141	0.011		0.949
Bad Reputation	1.986	0.352	168	0.043	0.949	
Other outcomes						
4. Days to decision, accepts only						
Good Reputation	1.643	0.144	220		0.083	0.001
No Reputation	2.368	0.451	157	0.083		0.023
Bad Reputation	3.943	0.499	196	0.001	0.023	
5. Promised wage rates						
Good Reputation	2.834	0.228	173		0.017	0.098
No Reputation	2.011	0.257	141	0.017		0.771
Bad Reputation	2.142	0.352	168	0.098	0.771	
6. Advertised pay						
Good Reputation	0.277	0.025	223		0.001	0.938
No Reputation	0.159	0.024	164	0.001		<0.001
Bad Reputation	0.28	0.022	211	0.938	<0.001	
7. RA log-seconds to complete						
Good Reputation	5.737	0.228	173		0.372	<0.001
No Reputation	5.639	0.085	141	0.372		0.001
Bad Reputation	6.368	0.069	168	<0.001	<0.001	

Note – Rejection rate p-values are from a χ^2 test that rejection rates are the same between the row and column. Time-to-pay p-values are from a two-sample t-test that the mean times-to-pay are the same between the row and column.

TABLE 2— Poisson regression for arrival of submitted tasks and other events

	Good Reputation		No Reputation			
Sample	β	SE	β	SE	periods	events
<u>Event: submitted tasks</u>						
Full sample						
(1) All submitted tasks	2.053*	(.132)	1.503*	(.102)	72	1641
Subsamples						
(2) Day 1 only	4.104*	(.467)	2.135*	(.264)	24	695
(3) Day 1-2 only	2.424*	(.196)	1.76*	(.15)	48	1125
(4) 12AM-6AM	1.679*	(.401)	1.393	(.345)	18	114
(5) 6AM-12PM	2.843*	(.35)	2.157*	(.277)	18	534
(6) 12PM-6PM	1.096	(.13)	.978	(.12)	18	415
(7) 6PM-12AM	2.694*	(.304)	1.648*	(.201)	18	577
Excluding last 12 hours						
(8) No controls	2.466*	(.185)	1.803*	(.142)	60	1313
(9) Controls for baseline rate	2.606*	(.201)	1.915*	(.156)	60	1313
(10) Day fixed effects	2.466*	(.185)	1.803*	(.142)	60	1313
(11) Hour fixed effects	2.093*	(.169)	1.471*	(.122)	60	1313
<u>Event: other</u>						
(12) Task previews	2.314*	(.142)	1.495*	(.099)	72	1837
(13) Task accepts	2.141*	(.133)	1.551*	(.102)	72	1799
(14) Error-free submissions	2.018*	(.165)	1.5*	(.129)	72	1012
(15) 1st submissions	2.871*	(.261)	1.644*	(.163)	72	899
(16) Error-free 1st submissions	2.88*	(.349)	1.641*	(.217)	72	508

NOTE – * $p < 0.05$. Each row is a regression. Coefficients are incident rate ratios with bad reputation as the omitted category. Standard errors in parentheses.

TABLE 3— Negative binomial regression for arrival of submitted tasks and other events

Sample	Good Reputation β	Good Reputation SE	No Reputation β	No Reputation SE	periods	events
Event: submitted tasks						
Full sample						
(1) All submitted tasks	2.053*	(.5)	1.503	(.368)	72	1641
Subsamples						
(2) Day 1 only	4.104*	(1.969)	2.135	(1.03)	24	695
(3) Day 1-2 only	2.424*	(.766)	1.76	(.559)	48	1125
(4) 12AM-6AM	1.679	(.823)	1.393	(.689)	18	114
(5) 6AM-12PM	2.843*	(1.201)	2.157	(.915)	18	534
(6) 12PM-6PM	1.096	(.267)	.978	(.239)	18	415
(7) 6PM-12AM	2.694*	(.955)	1.648	(.589)	18	577
Excluding last 12 hours						
(8) No controls	2.466*	(.704)	1.803*	(.516)	60	1313
(9) Controls for baseline rate	2.523*	(.719)	1.808*	(.515)	60	1313
(10) Day fixed effects	2.294*	(.654)	1.778*	(.498)	60	1313
(11) Hour fixed effects	1.858*	(.274)	1.374*	(.205)	60	1313
Event: other						
(12) Task previews	2.314*	(.571)	1.495	(.37)	72	1837
(13) Task accepts	2.141*	(.529)	1.551	(.384)	72	1799
(14) Error-free submissions	2.018*	(.548)	1.5	(.41)	72	1012
(15) 1st submissions	2.871*	(.804)	1.644	(.465)	72	899
(16) Error-free 1st submissions	2.88*	(.928)	1.641	(.536)	72	508

NOTE – * $p < 0.05$. Each row is a regression. Coefficients are incident rate ratios with bad reputation as the omitted category. Standard errors in parentheses.

Table 4— Preferred specification: negative binomial regression of arrival rates in the first sixty hours

	Previews (1)	Acceptances (2)	Submissions (3)	1st submissions (4)	Correct 1st submissions (5)
Good reputation	1.964*	1.909*	1.836*	2.488*	1.855*
	(0.280)	(0.277)	(0.262)	(0.426)	(0.405)
No reputation	1.403*	1.387*	1.364*	1.608*	1.261
	(0.204)	(0.203)	(0.196)	(0.277)	(0.278)
Constant	16.56*	14.10*	13.31*	8.024*	3.54*
	(4.907)	(4.300)	(4.002)	(2.788)	(1.729)
Day FE	Yes	Yes	Yes	Yes	Yes
Hour FE	Yes	Yes	Yes	Yes	Yes
Observations	60	60	60	60	60

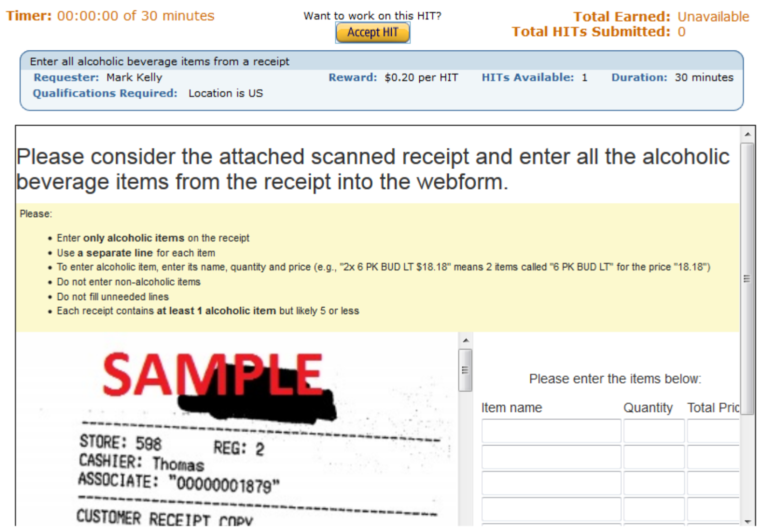
Note – *p< 0.05. Standard errors in parentheses. Bad reputation is the omitted category. All coefficients for good employers are significantly different from coefficients for bad employers with $p < 0.05$.

FIGURE 1: M-Turk worker's job search process: Turkopticon

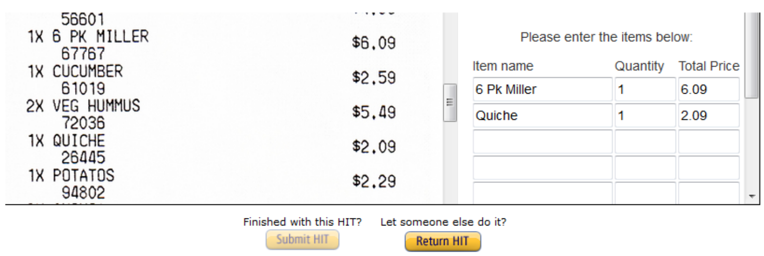
NOTE – Screen capture of a M-Turk worker's job search interface. The tooltip box left-of-center is available to workers who have installed Turkopticon, and shows color-coded ratings of the employer's communicativity, generosity, fairness, and promptness. It also offers a link to longform reviews.

FIGURE 2: M-Turk worker's job search process: previewing, accepting, and submitting tasks

Step 3: Workers preview tasks

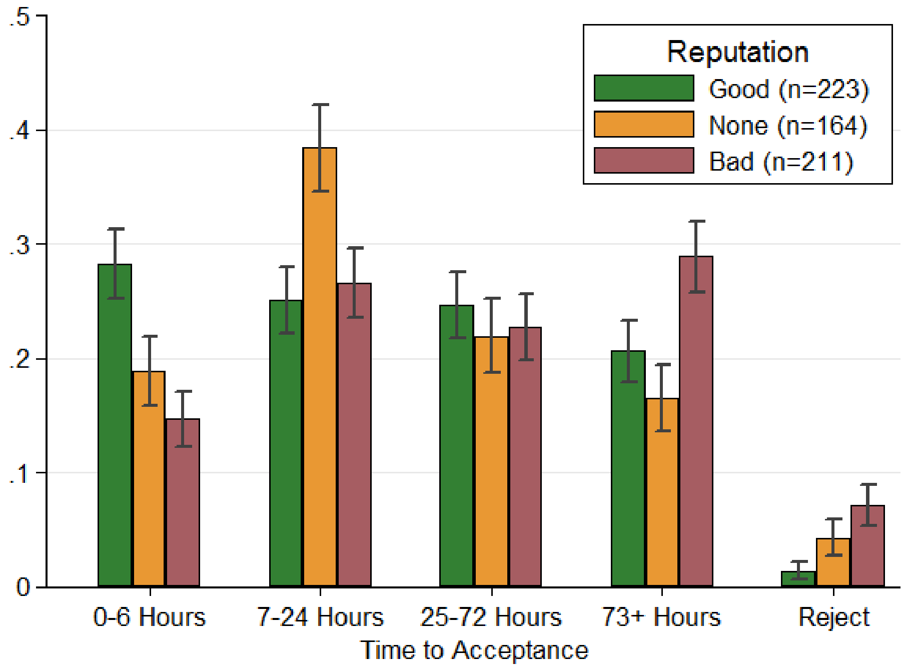


Step 4: Workers accept, perform, and submit tasks



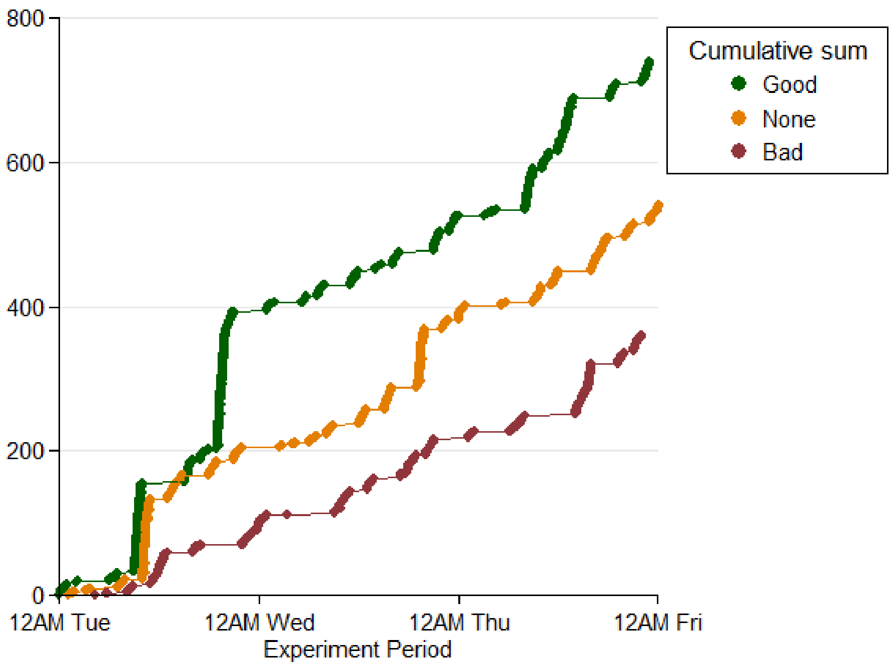
NOTE – Screen capture of a M-Turk worker's job search interface. From the list of tasks, workers must choose to preview a task before accepting the task. They then enter data into the webform and submit their work.

FIGURE 3: Time to payment and rejection by employer reputation



NOTE – Whiskers represent standard errors. p-values for a χ^2 test that shares are independent of reputation are respectively: 0.002, 0.011, 0.805, 0.012, and 0.007.

FIGURE 4: Cumulative accepted jobs by employer reputation



NOTE – Bold lines represent active job listings.

VII Appendix

FIGURE 5: Balanced, random allocation of employer identities to time-slots with reputation

	Tuesday	Wednesday	Thursday
0:00	Mark Kelly	Thomas Jordan	Mark Jordan
1:00	Joseph Warren	Joseph Jordan	Mark Warren
2:00	Thomas Warren	Mark Jordan	Joseph Kelly
3:00	Thomas Kelly	Thomas Jordan	Thomas Warren
4:00	Mark Warren	Joseph Warren	Mark Kelly
5:00	Joseph Kelly	Joseph Jordan	Thomas Kelly
6:00	Joseph Lewis	Thomas Lewis	Mark Lewis
7:00	Mark Roberts	Thomas Roberts	Thomas Clark
8:00	Thomas Clark	Thomas Lewis	Mark Clark
9:00	Mark Clark	Mark Lewis	Joseph Clark
10:00	Joseph Clark	Joseph Roberts	Joseph Lewis
11:00	Joseph Roberts	Thomas Roberts	Mark Roberts
12:00	Thomas Martin	Joseph Johnson	Joseph Martin
13:00	Thomas Adams	Joseph Adams	Mark Adams
14:00	Mark Martin	Mark Adams	Mark Johnson
15:00	Thomas Johnson	Thomas Adams	Joseph Adams
16:00	Mark Johnson	Thomas Johnson	Mark Martin
17:00	Joseph Martin	Thomas Martin	Joseph Johnson
18:00	Thomas Miller	Joseph Robinson	Thomas Robinson
19:00	Thomas Robinson	Mark Robinson	Thomas Owens
20:00	Mark Owens	Joseph Robinson	Mark Robinson
21:00	Joseph Owens	Joseph Miller	Mark Miller
22:00	Mark Miller	Thomas Miller	Joseph Miller
23:00	Thomas Owens	Mark Owens	Joseph Owens

NOTE – Red, green, and white denote employers endowed with bad, good, and no reputation, respectively.