

Scoundrels or Stars?
Theory and Evidence on the Quality of Workers in Online Labor Markets

Anne M. Farrell *
Miami University

Jonathan H. Grenier
Miami University

Justin Leiby
University of Florida

August 2014

* Corresponding author's contact information:

Farmer School of Business
Miami University
800 East High Street / 3094 Department of Accountancy / MSC 1002
Oxford, OH 45056
Phone: 513-529-1869
E-mail: Anne.Farrell@MiamiOH.edu

We are grateful for helpful feedback from workshop participants at Miami University and the University of Kentucky. We thank Garren Wood for programming assistance.

Scoundrels or Stars?
Theory and Evidence on the Quality of Workers in Online Labor Markets

ABSTRACT

Online labor markets give individuals and organizations the ability to quickly recruit large numbers of workers who are willing to work for relatively low wages. However, purchasers of labor in these markets confront severe adverse selection issues—including uncertainty about workers' honesty and willingness to exert effort. We conduct three experiments that examine how the honesty, effort allocation, and underlying motivations of online workers compare to other populations of workers. We find that online workers have honesty preferences similar to those found in prior research. We also find that online workers exert effort at least equal to, and sometimes greater than, those found in prior research. These effort levels are sustained even when pay is a flat wage that is significantly smaller than the average wage in prior research. Further, online workers increase effort in response to performance-based pay schemes only when the pay level is relatively low. Finally, we provide evidence that performance differences between online workers and traditional research participants is driven by online workers' relatively higher intrinsic motivation. We discuss implications for contracting with online workers and for the use of online workers as participants in accounting research.

Keywords: Mechanical Turk; online labor markets; honesty; effort; performance-based pay; intrinsic motivation

I. INTRODUCTION

Online labor markets are a growing means for organizations and individuals to find and hire workers with particular skills to complete a variety of tasks, including office work, commercial or web design, marketing, writing, translation, accounting, finance, programming, and software testing (e.g., oDesk; uTest, now Applause; Guru.com; Elance; 99designs; CrowdFlower).

Amazon.com's Mechanical Turk ("MTurk") is one such site, where workers are paid relatively small sums to complete simple tasks such as extracting information from scanned documents, writing captions for photos, or validating web sites or phone numbers. While some online labor markets focus on workers with specialized skills, others like MTurk include diverse workers willing to work for very low pay (perhaps as low as \$1.38 per hour; see Horton and Chilton 2010). The large population and low labor costs of online labor markets are appealing to purchasers of labor.

Yet, purchasers of online labor confront severe adverse selection issues. For instance, workers on MTurk are anonymous and often self-report their attributes. Because purchasers are uncertain *ex ante* whether the workers they hire are high or low quality, economic theory predicts that purchasers are likely to offer wages between the wages of high quality workers and those of low quality workers. In turn, high quality workers are likely to exit the labor pool, causing wages and labor quality to spiral downwards. In this rationale, the low wages of some online labor markets reflect the low quality of the labor pool. Research has not yet addressed these fundamental issues in contracting in online labor markets.

In this paper, we examine the quality of labor in online markets as it relates to whether online workers are honest in their reporting to principals, exert reasonable amounts of effort given relatively low wages, and are sensitive to differences in compensation contract design.

These issues are of particular interest in contracting settings characterized by high information asymmetries, complex and often lengthy tasks, and variations in pay schemes. Because many inferences in the contracting literature are based on experiments with student participants, we use the performance of students in representative studies from prior literature as a benchmark for online workers' performance. Further, we examine differences in the motivations of online workers relative to more traditional populations. We propose that online workers have different motivations than do other workers—including different concerns about reputation and different degrees of intrinsic enjoyment of computer tasks.

With respect to honesty preferences, online workers' reputations are important for securing more profitable and interesting tasks in the market, and some research suggests they are unlikely to differ from populations examined in prior research (Paolacci, Chandler and Ipeirotis 2010; Buhrmester, Kwang and Gosling 2011; Horton, Rand and Zeckhauser 2011; Goodman, Cryder and Cheema 2012). On the other hand, pay in many online markets is notoriously low, and low wages have been found to increase dishonest behavior by workers (Chen and Sandino 2012). Further, online interactions produce lower preferences for fairness and reciprocity and a lower sense of accountability than do face-to-face interactions ((Bohnet and Frey 1999; Valley, Moag and Bazerman 1998; Kachelmeier and Towry 2002; Brazel, Agoglia and Hatfield 2004). We predict that the anonymity and weaker social norms in online labor markets will result in online workers having lower preferences for honesty over wealth than those of other populations that have been found in prior research.

With respect to effort, online workers' low reservation wage may reflect their underlying poor quality. Also, low wages and weak monitoring may not provide sufficiently strong motivation to induce significant effort (Paolacci et al. 2010; Goodman et al. 2012; Mason and

Suri 2012). On the other hand, online workers' may derive enjoyment simply from completing online tasks and their reputations affect their ability to secure desirable work. Thus, if intrinsic enjoyment induces sufficient motivation for more complex, time-consuming, but still inherently interesting tasks, online workers may exert equally significant levels of effort and be less sensitive to differences in performance-based contracts relative to findings of prior research.

We conduct three studies to examine these issues. In studies 1 and 2, we chose a validated task from prior research that requires greater attention and effort than many tasks found in online labor markets (Evans et al. 2001; Farrell et al. 2008). This allows us to benchmark online workers against another population in a setting in which online workers' low quality is theoretically most likely to manifest. Further, for each study, we include contracts with wage levels that are identical to the original studies in addition to contracts with wages that are five orders of magnitude smaller. This allows us to address the cost-effectiveness of online labor markets relative to other potential populations. In study 3, we administer a variation of the task from Farrell et al. (2008) to both students and online workers, to enable us to compare the underlying motivations of the two groups.

In our first study, MTurk workers complete the managerial reporting task from Evans et al. (2001). Over ten rounds, participants act as agents who privately observe cost signals and determine whether they will report these signals truthfully to the principal, knowing that they can keep any surplus gained from misrepresentation for themselves. Our results indicate that online workers are just as honest, if not more so, than participants in prior research, regardless of pay level. Based on these results, we conclude that online workers have similar preferences for honesty over wealth as those reported in prior research.

In our second study, workers in MTurk complete the multi-period sandwich making task

from Farrell et al. (2008), which captures both the quantity and quality of effort. We manipulate the terms of two contracts as a flat wage contract and a performance-based variable wage contract to motivate different effort allocations. We selected the sandwich making task in part because it is likely to be enjoyable, particularly to those who enjoy computer tasks (Paolacci et al. 2010).

Contrary to the idea that online workers are effort averse, we find that effort under both performance-based and flat wage contracts is as good as or better than that of participants in Farrell et al. (2008), even when the magnitude of pay is five times smaller. In fact, online workers who received only a \$5 flat wage for the hour-long task exert effort levels similar to those of performance-compensated participants whose average pay was about \$25 in Farrell et al. (2008). With respect to sensitivity of effort to differences in the terms of performance-based contracts, we find that performance-based contracts increase effort only when the pay level is relatively low. When pay is relatively high, effort levels are invariant and high across contract types. These results are consistent with the idea that online workers may derive substantial enjoyment from completing computer tasks.

In our third study, online workers and student participants complete a modification of the Farrell et al. (2008) study that we designed to be substantially less game-like than study 2. If online workers' relatively high intrinsic motivation on game-like tasks caused their performance to be (1) equivalent to that of lab participants and/or (2) insensitive to performance incentives, then either of these findings may differ in a less enjoyable task. We again manipulate contract type as a variable wage contract as opposed to a flat wage, and include low and high pay levels as in the first two studies. As in study 2, we find that online workers' performance is equivalent to students'. We also find only qualified evidence of pay-performance sensitivity among online

workers—specifically, differences in intrinsic enjoyment and reputation concerns. A mediation analysis reveals that performance differences are largely attributable to the relatively higher intrinsic enjoyment of online workers.

This paper makes an important contribution to the understanding of contracting in online labor markets. In brief, we address a fundamental concern about adverse selection in online labor markets, which compromises the potential for efficient contracting in such markets. Our findings suggest that the quality of labor in online markets may be higher than theory would predict, given the features of the market. As online labor markets become more popular for completing work, it is important to understand how online labor markets compare to other sources of labor.

Also, our study offers several contributions of interest to accounting researchers in multiple areas. Our findings indicate that researchers interested in accounting domains in which honesty preferences are an important or assumed construct—for example, tax evasion, financial reporting fraud—can comfortably consider using online workers in their studies. Similarly, accounting researchers can also comfortably consider using online labor markets for settings in which high effort expenditures or high intrinsic motivation are required. Our results indicate that comparable or greater effort can be induced at substantially lower wages than those paid in prior research, although certainly fair wages should be paid for the work requested. As such, we contribute to discussions about the suitability of particular participant pools for various research populations of interest by identifying a potentially cost-effective, quick source for reliable data (e.g., Peecher and Solomon 2001; Libby, Bloomfield and Nelson 2002; Elliott, Hodge, Kennedy and Pronk 2007; Henrich, Heine and Norenzayan 2010).

II. BACKGROUND AND HYPOTHESES

Overview of Online Labor Markets

Online labor markets have become a popular tool for individuals and organizations to quickly identify short-term workers who have skills required for particular, often one-time, tasks. Their popularity is indicated not only by their proliferation but by numerous commentaries and articles in the popular press, including the Wall Street Journal (Holmes 2009; Glazer 2011; Silverman 2012), the New York Times (Pontin 2007; Pattison 2009; Folbre 2013), Forbes (Worstall 2013), and the Huffington Post (Dobson 2013).

Amazon's Mechanical Turk is one such market. MTurk enables "Requesters" to post "HITs" (human intelligence tasks) that workers can browse and choose to complete for pay. Tasks listed on MTurk are those that cannot be completed by computers, such as listing items on scanned documents, identifying performers on music clips, describing photographs, or verifying phone numbers or web addresses. MTurk reportedly includes more than 500,000 workers in 190 countries (Ross et al. 2010; Amazon.com 2014a); in 2013 the approximate number of HITs available on a given day ranged from 60,000 to 475,000 (MTurk Tracker 2014).

To establish MTurk accounts, both Requesters and workers provide Amazon with a legally-valid name and contact information. However, Requester and worker identities are typically anonymous and cannot be ascertained by user names in the market.

Requesters fund a prepaid account with Amazon before posting HITs. Each posted HIT includes a task preview; the time limit for completing the HIT once it is accepted; the pay rate and possible bonus (if any); how many comparable HITs from this Requester are available; and how long the HITs will be available in the market. Requesters can also screen workers by establishing "Qualifications", or standards that must be met before accepting the HIT (e.g., location of residence; passing a test; have not previously completed tasks for the Requester).

Workers must provide Amazon with tax identification information, and acknowledge that

they have only one worker account and will not use automated means to complete tasks. Once a worker qualifies for and accepts a HIT, s/he must complete it within the predetermined amount of time or it is released back into the market. Once the task is completed, the Requester reviews and either approves or rejects the worker's work. If the HIT is approved, Amazon uses the funds in the Requester's prepaid account to disburse pay to the worker.¹

Amazon does not report average pay rates across MTurk tasks, and estimates by outsiders vary. Some blogs suggest that workers can earn from \$8 to \$16 per hour by completing many tasks (e.g., Chen 2012; Roggio 2012), while both academic literature and media reports suggest wages of \$1.20 to \$5.00 per hour, or even less (e.g., Ross et al. 2010; Buhrmester, Kwang and Gosling 2011; Dobson 2013; Folbre 2013). An analytical model and experiment reports a median reservation wage of \$1.38 per hour (Horton and Chilton 2010). Regardless of pay rate, reputations matter. Amazon computes workers' "Approval Rate" and "Abandoned HITs" (i.e., HITs not completed), and bestows a title of "Master" on high-performing workers (Amazon.com 2014b). Requesters can use these measures to screen workers, and workers whose work on HITs are rejected have been known to confront Requesters via e-mail. At the same time, forums and web sites like *TurkerNation* and *TurkOpticon* allow workers to rate and review Requesters.

Research on Online Labor Markets

The use of online labor markets is increasingly popular. In particular, the use of MTurk to recruit research participants has become popular in a variety of fields (see Chandler, Mueller and Paolacci 2014 for examples). Advantages of online markets include their large, diverse pools of workers who complete tasks quickly, around the clock, for pay that is often significantly lower than that of other groups; the simple infrastructures for posting tasks and recruiting and paying

¹ Amazon receives a 10%-20% commission on each approved HIT. U.S.-based Turkers can receive pay via either a bank transfer or an Amazon gift certificate. India-based Turkers can receive pay via either a check in Indian Rupees or an Amazon gift certificate. Turkers in all other countries can only receive pay via an Amazon gift certificate.

participants; and the anonymity that limits the abilities of researchers to influence responses and of participants to communicate about task details (Paolacci et al. 2010; Mason and Suri 2012; Crump, McDonnell, and Gureckis 2013). On the other hand, disadvantages of online markets include the aforementioned adverse selection issues, as well as uncertainties about whether such samples mirror demographic characteristics of populations of interest; whether workers do in fact communicate with each other about task details; and the extent to which workers report honestly given their anonymity. Further, online workers work for low pay on often mundane tasks, and potentially work on multiple tasks simultaneously or in distracting environments. Thus, questions arise about their attention and effort (Paolacci et al. 2010; Buhrmester et al. 2011; Suri, Goldstein, and Mason 2011; Goodman et al. 2012; Chandler et al. 2014). Recent research has begun to address some of these uncertainties with respect to MTurk.

With respect to demographics (Table 1), about half of the online workers on MTurk surveyed by Paolacci et al. (2010) and Ross et al. (2010) are U.S.-based. Compared to the U.S. population, Paolacci et al. (2010) report that average age and income levels of U.S.-based workers on MTurk are slightly below U.S. averages; education level is slightly higher; and the proportion of females is higher. Buhrmester et al. (2011) compare workers on MTurk, other internet samples, and U.S. college samples, and find that workers on MTurk are more diverse than both groups. Paolacci et al. (2010) report that only 14 percent of workers on MTurk used MTurk as their primary income source; 61 percent state that earning money is an important motivation for their work; and 41 percent participate in the market for entertainment.

----- Insert Table 1 about here -----

With respect to communications and honesty, Chandler et al. (2014) examine whether workers on MTurk communicate with each other about research tasks or complete a HIT

multiple times. They find that online worker cross-talk focuses more on HIT length and payment rather than content, and that online worker forums have implicit and explicit norms against sharing HIT content. Regarding research, Chandler et al. (2014) find that some online workers are in general more likely to choose to complete research studies, so online worker samples on such HITs may not be representative of the online worker population. Finally, Goodman et al. (2012) find that online workers are more likely to look up correct answers on quizzes even when told not to do so, and Suri et al. (2011, 66) report that on die-roll tasks, “while few [online workers] cheated a lot, many cheated a little”.

Finally, evidence on online workers’ effort and work quality is mixed. Paolacci et al. (2010) find that online workers are no more likely than laboratory participants to fail an attention check question. Conversely, Goodman et al. (2012) find that online workers pay less attention than laboratory participants to experimental materials, and Crump et al. (2013) find that including online workers who fail attention checks in analyses introduces substantial noise in results. In these studies, however, online workers are at least similar to, if not the same as, laboratory participants with respect to performance on cognitive tests (e.g., Stroop), risk preferences, and susceptibility to heuristics and biases (e.g., framing effects, outcome bias; see also Horton et al. 2011).

In summary, research has begun to examine characteristics of workers in online labor markets to enable researchers to judge whether they are appropriate for their research questions.² Beyond this methodological issue, more fundamental questions remain unanswered about contracting with online workers. For example, questions still remain about their honesty in the face of information asymmetries that are common in business settings; their effort on longer and more

² Several prior studies have also provided guidance on the mechanics and sound conduct of research on MTurk (e.g., Paolacci et al. 2010; Buhrmester et al. 2011; Horton et al. 2011; Mason and Suri 2012; Chandler et al. 2014).

complex tasks; and the sensitivity of that effort to different compensation contracts, also a common feature of business settings. Indeed, people who are dishonest and willing to work for relatively small amounts of pay may not be representative of the broader population. At the same time, it is difficult to reconcile how online workers could be simultaneously effort-averse *and* willing to work for a pittance. In the next section, we examine theoretical and practical reasons why online workers' honesty, effort, and pay-performance sensitivity may or may not be similar to that found in prior research.

Online Workers' Honesty Preferences

Honesty in communication in the face of information asymmetries is crucial to understanding the economic activities of organizations and markets. Early economic theory suggests that the opportunistic behavior of agents limits honesty (e.g., Baiman and Evans 1983; Penno 1984; Melamad and Reichelstein 1987). However, later work recognizes that agents have other preferences that may limit the extent of this behavior (e.g., Fehr and Schmidt 1999; Evans et al. 2001; Fehr and Falk 2002; Fehr and Fischbacher 2002; Hannan et al. 2006; Rankin, Schwartz, and Young 2008).

In their seminal study, Evans et al. (2001) had MBA students act as division managers with private information about production costs. Over multiple periods, their task was to report production costs to headquarters, and they kept the difference between reported and actual costs. Despite the opportunity and incentives to report dishonestly, participants exhibited substantial preferences for honesty over wealth that did not change as payoffs for dishonesty increased. However, the optimal contract for balancing the benefits of limiting agents' ability to lie against the lost profits from no production (which provided participants with a smaller share of profit) resulted in weaker preferences for honesty.

Honesty has continued to receive considerable attention in accounting since Evans et al. (2001), with a particular interest in understanding the contexts in which preferences for honesty over wealth vary. Hannan et al. (2006) find that relatively more precise information systems increase dishonesty by magnifying the benefits of over-reporting. Social preferences also have been found to influence honesty. For instance, preferences for honesty strengthen as the perceived fairness of the principal increases (Zhang 2008). Church et al. (2012) find that dishonesty increases when its benefits, such as budget slack, are shared with others. Maas and van Rinsum (2013) find that dishonesty increases when it benefits others, but decreases when it hurts others.

Given the importance of honesty to the functioning of firms, markets, and governments and the continuing interest in better understanding how context affects honesty preferences, the honesty construct is likely to continue to be important in contracting research. Thus, it is important to understand the implications of using online labor markets for tasks that involve honesty or assume preferences for honesty over wealth.

Some evidence suggests that online workers' honesty preferences may mirror patterns found in prior research. First, online workers appear to be demographically similar to the population as a whole, albeit more diverse than students who were often the participants in prior research (Paolacci et al. 2010; Buhrmester et al. 2011). Second, online workers exhibit many of the same preferences and susceptibilities to heuristics and biases found in prior literature (see, e.g., Paolacci et al. 2010; Horton et al. 2011; Goodman et al. 2012). Third, as noted earlier, online workers' reputations in the labor market affect their ability to secure desirable work.

On the other hand, several distinguishing features of online labor markets and their workers suggest that online workers may have lower preferences for honesty than observed in prior

research. First, online workers are often willing to work for relatively low pay, which leaves open the possibility that they also have relatively low thresholds to engage in dishonest behavior. Second, Goodman et al. (2012) find that online workers are willing to work for small incentives because they value their time less than the broader population, not because they value money less; in fact, online workers may be more materialistic than the population at large. Third, Goodman et al. (2012) also find that online workers take advantage of their unsupervised environments to look up correct answers on tasks that require them.

Fourth, online workers do not have any social contact with researchers, while lab participants may at least be familiar with the researcher's name since they are often in the same organization. Online workers are in different locations, while lab participants are typically in the same town or even the same room as the researcher. Finally, given anonymity, online workers may perceive that the chance and consequences of being caught in a lie are remote and minor, compared to the perceptions of lab participants. Indeed, prior research finds that social preferences for fairness and reciprocity toward peers are lower in remote rather than in-person interactions (Bohnet and Frey 1999; Valley et al. 1998; Kachelmeier and Towry 2002), and posits that these differences may be due to the relative absence of social information and weakened social norms (Hoffman, McCabe and Smith 1996; Bazerman, Curhan, Moore and Valley 2000). Further, Brazel et al. (2004) find that workers (in this case, auditors) feel more accountable to their supervisors in face-to-face than in electronic interactions.

Given these distinguishing characteristics of online labor markets compared to the more traditional lab settings used in prior research, our first hypothesis predicts that online workers have weaker preferences for honesty than have been found in prior research.

H1: Workers in online labor markets have weaker preferences for honesty than have been found in prior research.

Online Workers' Effort

Even if accounting research questions are amenable to testing with a broad, diverse population, the tasks in the studies can be complex and time-consuming. Because online workers appear to value their time less than does the population at large (Goodman et al. 2012), these labor markets may be a resource for conducting accounting research at substantial savings. At the same time, given their greater complexity, tasks in accounting studies require sufficient and steady attention to details and the motivation to exert effort (Sprinkle 2003).

Online workers do have reputation concerns in their markets that may in and of themselves induce high attention and effort levels. Further, many online workers report that they are motivated to complete tasks simply for entertainment (Paolacci et al. 2010). However, researchers' inability to supervise online workers, coupled with the workers' ability to work on multiple tasks or even walk away, lead to serious concerns about online workers' attention and effort (Oppenheimer, Meyvis and Davidenko 2009; Mason and Suri 2012). These concerns are compounded by the low pay rates and mundane tasks in many online labor markets, which calls into question whether workers attracted to these markets have the ability to complete more complex tasks in the first place.³

There is mixed evidence on online workers' attention. Goodman et al. (2012) find that online workers are more likely than student participants to fail attention checks. Conversely, Paolacci et al. (2010) find no differences in attention between samples of online workers and undergraduates. Regarding effort, Goodman et al. (2012) find that online workers perform worse

³ Given that some workers in online labor markets participate purely for entertainment, there is emerging evidence that "gamers" in general do not necessarily deserve the negative reputation the word connotes. Specifically, 69 percent of all heads of household play computer and video games; 25 percent of all "gamers" are over the age of 50; the average "gamer" is 35 years old and has been playing games for 12 years; and 61 percent of surveyed CEOs, CFOs, and other senior executives say they take daily game breaks at work (McGonigal 2011, p. 11).

than undergraduates on a task that measures cognitive effort, while Crump et al. (2013) find that online workers perform similarly to students on cognitive tests such as Stroop and task-switching. Notably, though, while these tasks require attention, they are less complex than those often used in accounting research.

As a result, it is an open and important question as to whether online workers are motivated to exert effort, and whether that varies with pay levels. Given the weight of evidence on online workers' attention and effort and the conditions under which they complete their tasks, we predict that online workers' effort levels are relatively low at standard pay levels for online markets. Further, we expect that even if pay levels are increased, effort will not reach the levels found in prior research given uncertainty about worker attention, motivation, and skills relative to other research populations – and offering higher pay negates a significant benefit of online labor markets.

H2: The effort levels of workers in online labor markets are lower than those found in prior research.

Pay-Performance Sensitivity of Online Workers' Effort

Many accounting research studies involve contexts in which workers' efforts are sensitive to compensation contract type. For example, studies of variations in incentive contracts require that the willingness to exert effort can be influenced by differences in the contracts themselves (see Young and Lewis 1995 for a review). In this regard, doubts exist about the sensitivity of online workers' efforts to differences in compensation contracts.

First, online workers may exert comparable effort levels regardless of contract type given their reputation concerns in the market. Second, many online workers report being motivated by factors other than money. Specifically, 41% of online workers in Paolacci et al. (2012) report that they complete tasks for entertainment purposes, suggesting a high level of intrinsic

motivation for computer-based tasks. While this may also be true of participant populations used in prior research (e.g., students), unlike workers in online labor markets, these participants typically also receive extrinsic incentives such as extra credit in courses and standard labor market wages. The fact that many online workers report intrinsic rewards from performing relatively mundane tasks raises the possibility that their motivation to exert effort is relatively stable, and thus insensitive to incentive contract differences (see Gneezy and Rustichini 2000, Dickinson and Villeval 2008 for discussions of “crowding out”).

We argue that this issue is of particular interest due to the proliferation of accounting experiments that use intrinsically interesting tasks. For example, recent studies have asked participants to create “rebus puzzles,” in which words, pictures, symbols, and numbers are used to depict familiar words or phrases (Kachelmeier et al. 2008; Kachelmeier and Williamson 2010), or to develop proposals for the use of a neglected historic building (Chen et al. 2012). Participants in Choi et al. (2012, 2013) compete in the life simulation computer game *Spore*. In Farrell et al. (2008, 2012), participants make virtual sandwiches with a game-like interface.

If online workers are, on average, intrinsically motivated, then it is likely that their efforts will be relatively insensitive to differences in compensation schemes. This is consistent with literature demonstrating that intrinsically-motivated individuals are unlikely to increase their efforts in response to performance-based incentives (e.g., Deci, Koestner and Ryan 1999; Ryan and Deci 2000). As such, we hypothesize that the efforts of workers in online labor markets are insensitive to differences in compensation contracts.

H3: The efforts of workers in online labor markets are insensitive to differences in compensation contracts.

III. RESEARCH DESIGN OVERVIEW

To test our hypotheses, in studies 1 and 2 we selected two experimental studies based on

their representativeness for our research question and the feasibility with which they could be administered using a readily-available online labor market already used in research, MTurk. Regarding representativeness, we selected studies that are archetypal and thus contained minimal idiosyncrasies, such as unusual participant pools. The feasibility standard precluded studies that would have required, for instance, the installation of specialized software or the physical presence of participants. Thus, to test our first hypothesis on honesty preferences (Study 1), we use the reporting task from Evans et al. (2001), which has been used in similar studies of honesty (e.g., Hannan et al. 2006; Rankin et al. 2008). To test our second and third hypotheses on effort and pay-performance sensitivity (Study 2), we use the production task from Farrell et al. (2008), which has been used in at least one subsequent study (Choi, Newman and Tafkov 2013). Both studies were approved by the institutional review boards at the authors' universities.⁴ For study 3, we use a modification of the Farrell et al. (2008) task that is described in greater detail in Section 6.

Participants

Participants in all three studies were recruited and paid using Amazon's Mechanical Turk online labor market. Due to strict policies at the researchers' universities that govern payments to non-U.S. citizens, we used a HIT qualification limiting participation to U.S. citizens. Further, we also qualified that participants could not have previously completed a HIT for the researcher, and checked for duplicate HITs using worker identification ID numbers and IP addresses. See Table 1 for demographic information for participants in both studies, including comparisons of their scores on a financial literacy quiz to those of business students and the population at large.

⁴ For studies 1 and 2, we use reported findings from prior studies as benchmarks, rather than repeating those studies today with the same participant populations. First, we have no reason to expect that preferences for honesty, effort, or pay-performance sensitivities would vary dramatically today from those reported in the original studies. Second, our approach conserves research participants and controls the costs of our study.

MTurk Procedures

For each study, we gathered data on the same task at two pay levels – high pay, or the level used in the original study, and low pay, or five times smaller than the pay level in the original study. We posted HITs on MTurk that included the pay rate; possible bonus (since both studies use performance-based contracts); estimated time to complete the task; and a limit of two hours to complete the HIT once accepted.

Data for the high and low pay versions of Study 1 was gathered in one day. Data for the high pay version of Study 2 was gathered over two days, and the low pay version over five days, since our web server could not host enough participants for that task in one day. We ran the low pay version of Study 2 first, followed by the high pay version one week later; we then ran Study 1 twelve days later. During this time, we monitored MTurk-focused blogs and discussion boards; while we did not observe cross-talk about the content of the tasks themselves, we did observe postings identifying Study 2 as fun and profitable. For study 3, data for online workers was collected using MTurk over the course of three days. Data from students was collected over the course of eight weeks.

IV. STUDY 1 – HONESTY

Research Design

Procedures

The study was designed and hosted using Qualtrics. Instructions and procedures were nearly identical to those in Evans et al. (2001); we note differences in the following description.

Participants were asked to assume they were division managers in a manufacturing firm, and that their task was to report production costs to corporate headquarters over a series of ten work periods. Both headquarters and the manager knew the distribution of possible costs, but the

manager privately observed actual costs. After observing each period's realization of actual cost, participants submitted a budget request to headquarters. They kept any surplus from submitting a budget request that was higher than actual cost, as well as a flat salary each period.

After participants read the instructions, we provided three examples of actual cost-budget request combinations with computations of participant earnings and firm profit. A comprehensive earnings spreadsheet was available as a pop-up window throughout the task; it detailed the earnings and firm profit implications of all possible actual cost-budget request combinations. Participants then began the ten-period task. When finished, participants answered post-task questions to insure the task was not completed by an automated "bot", and about their understanding of and beliefs about the task and their backgrounds.

As in Evans et al. (2001), participants were informed that their cost reports were anonymous. In a post-task question, nine participants indicated that they did not believe that their responses were anonymous; since inclusion or exclusion of their cost reports does not change any of our inferences, we include them for completeness. Participants were also informed that their compensation would be based on one randomly-determined round, so they had no incentives to alter their reporting behavior from period to period.

Independent Variables

We ran two experiments that varied only by the conversion rate of Lira to dollars. In the *high pay* experiment, participants were compensated at a rate of 30 Lira = \$1, the same conversion rate used in Experiments 1 and 3 of Evans et al. (2001).⁶ In the *low pay* experiment, participants were compensated at a rate of 150 lira = \$1. Thus, possible earnings in the *high pay* experiment were equivalent to prior studies, while possible earnings in the *low pay* experiment

⁶ We do not use the high-payoff trust contract in Experiment 2 of Evans et al. (2001) because our *high pay* experiment already provides payoffs that are significantly higher than normal in the MTurk labor market.

were substantially smaller but more representative of pay levels in MTurk tasks.

Within each experiment, we manipulated *contract type* between subjects. Following Evans et al. (2001), participants compensated under the *trust contract* could report any amount between 4.00 and 6.00 Lira per unit and keep the surplus above actual cost. Under the *modified trust contract*, a production hurdle was incorporated such that the budget request would be denied if the reported cost was above 5.00 Lira. Managers receive their requested amount if it is less than or equal to 5.00 Lira per unit, but only their salary if their request exceeds 5.00 Lira.

Results

Attention, Comprehension, and Manipulation Checks

As an attention check, we asked participants to answer “true,” “false,” or “don’t know” to the statement, “Your task in the experiment was to submit a budget request each period.” Of the 128 participants across the *low* and *high pay* experiments, 127 (99.2%) answered this correctly. Excluding this participant does not affect our inferences.

We also assessed participants’ attention with a series of six true/false questions to measure their understanding of the task.⁷ We coded each incorrect response as one and summed these measures to create a *misunderstanding* score, with a range of 0 (no misunderstanding) to 6 (complete misunderstanding). Mean (standard deviation of) *misunderstanding* was 0.77 (1.05) across both experiments. Thus, participants appeared to have attended to the instructions and understood the task.

Dependent Variable

Following Evans et al., honesty was measured with *percentage of honesty*, computed as:

⁷ Amongst other questions, these included, “Your personal earnings for one period in the experiment affected your earnings in other periods of the experiment”, and, “Your personal earnings for any period in the experiment depended on the amount of your budget request”.

$$1 - \left(\sum_{i=1}^n \text{Payoff Claimed} / \sum_{i=1}^n \text{Payoff Available} \right)$$

This measures the percentage of the total available payoff across periods that a participant did not claim.⁸ As in Evans et al. (2001), if a participant requested an amount below actual cost in a period, we replace the report with the actual cost; this decreases the percentage of honesty and firm profit, so our honesty measure is not inflated.

Hypothesis 1 Results – Honesty

Recall that H1 predicts that online workers will have weaker preferences for honesty than have been found in prior research. We use the *percentage of honesty* results in Evans et al. (2001, Table 1) as benchmarks against which to compare the honesty of Online workers. Across both *high* and *low pay* experiments and both contract type conditions, *percentage of honesty* was 38.2%. Table 2 reports statistics for *contract types* in the *low* and *high pay* experiments.

----- Insert Table 2 about here -----

First, we compare online workers' preference for honesty to those reported in Evans et al. (2001). Under the *trust contract*, the *percentage of honesty* in Evans et al. (2001) does not statistically differ from online workers' honesty with either *high pay* ($M_{\text{EvansTrust}}=48.7\%$ versus $M_{\text{TurkTrustHigh}}=57.3\%$, $t(22)=1.46$, $p=0.16$) or *low pay* ($M_{\text{EvansTrust}}=48.7\%$ versus $M_{\text{TurkTrustLow}}=49.9\%$, $t(46)=0.24$, $p=0.81$). Likewise, under the *modified trust contract*, the *percentage of honesty* in Evans et al. (2001) does not statistically differ from online workers' honesty with either *high pay* ($M_{\text{EvansModified}}=21.8\%$ versus $M_{\text{TurkModifiedHigh}}=15.3\%$, $t(18)=-0.96$,

⁸ This example is taken directly from Evans et al. (2001, 541): For an actual cost draw of 5.00 Lira per unit, the *payoff available* to the manager was 1,000 Lira. That is, by reporting 6.00 Lira per unit the manager would receive 6,000 Lira from the firm to produce 1,000 units, of which 5,000 Lira would be used for production, yielding a payoff available of 1,000 Lira. If instead the subject reported 5.60 Lira per unit, the firm would transfer 5,600 Lira to the manager. The manager would use 5,000 Lira for production, yielding a *payoff claimed* of 5,600 - 5,000 = 600 Lira. The percentage of honesty (π) would equal $[1 - (\text{payoff claimed} / \text{payoff available})] = [1 - (600 / 1,000)] = 40\%$. If the subject reported 5.00 Lira per unit (i.e., reported completely honestly), then π would equal 100%; and if the subject reported 6.00 Lira per unit (reported to maximize wealth), π would equal 0%.

$p=0.35$) or *low pay* ($M_{\text{EvansModified}}=21.8\%$ versus $M_{\text{TurkModifiedLow}}=24.0\%$, $t(38)=0.43$, $p=0.67$).

These findings indicate that online workers' preferences for honesty are equivalent to that of the MBA student participants in Evans et al. (2001). Thus, H1 is not supported. online workers' relatively low reservation wage does not influence their reporting honesty, nor does it lead them to report less honestly as the payoffs to doing so increase.

Second, we compare preferences for honesty amongst online workers across the two contract types. In the *high pay* experiment, *percentage of honesty* was greater with the *trust contract* than the *modified trust contract* ($M_{\text{TurkTrustHigh}}=57.3\%$ versus $M_{\text{TurkModifiedHigh}}=15.3\%$, $t(40)=4.76$, $p<0.01$). Likewise, in the *low pay* experiment, *percentage of honesty* was greater with the *trust contract* than the *modified trust contract* ($M_{\text{TurkTrustLow}}=49.9\%$ versus $M_{\text{TurkModifiedLow}}=24.0\%$, $t(84)=3.75$, $p<0.01$). This pattern is consistent with that in Evans et al. (2001) ($M_{\text{EvansTrust}}=48.7\%$ versus $M_{\text{EvansModified}}=21.8\%$).

Third, we examine the pattern of fully dishonest reports – i.e., participants who claimed the full available payoff in every round and, thus, strictly conformed to the predictions of economic theory. The pattern for online workers also mirrors that in Evans et al. (2001). While few reports were fully dishonest in the *trust contract* conditions (15% and 0%, in the *low pay* and *high pay* experiments, respectively, compared to 27% in Evans et al. 2001), roughly half of the reports in the *modified trust contract* conditions were fully dishonest (49% and 53% in the *low pay* and *high pay* experiments, respectively, compared to 61% in Evans et al. 2001).

In sum, it is noteworthy that the findings of Evans et al. (2001) were replicated with online workers. First, with the high pay experiment, the potential payoffs from dishonesty were several orders of magnitude greater than the average payoffs that MTurk workers typically earn. Second, with the low pay experiment, online workers do not appear to have a different threshold for

dishonest reporting even though their wages are typically low. Thus, despite the anonymity and remote, unsupervised environments of online labor markets, online workers appear to have preferences for honesty that are roughly equivalent to those found in prior research. Further, these workers responded to differences in compensation contract structures in a manner consistent with that observed in prior studies.

V. STUDY 2 – EFFORT AND PAY-PERFORMANCE SENSITIVITY

Research Design

Procedures

The study was designed using custom software and hosted on a private web site. Instructions and procedures were nearly identical to those in Farrell et al. (2008); we note differences in the following description.

Participants were asked to assume that they were employees fulfilling customer orders in a sandwich shop for multiple four-minute work periods using a computerized order and assembly system. Prior to starting the actual work periods, participants received detailed instructions on the task and completed a practice round. The program included 50 unique ingredient combinations, and sandwich orders appeared in a pre-programmed order. When a customer's order appeared, participants used a drop-down menu to find the sandwich on the menu of all sandwiches, and reviewed that sandwich's ingredient list. The ingredient list disappeared if the mouse was moved away from the menu, but participants could return to review the ingredient list as often as desired. Participants then assembled the sandwich by selecting ingredients from five drop-down menus, each with a different category of ingredients. After clicking a "finished" button, the computer checked the assembled sandwich against the ingredient list and tallied the number of mistakes in the assembled sandwich. If the sandwich had four or more mistakes, the

order had to be remade. The next customer order then appeared on the screen.

After the practice round, participants were told that the sandwich shop's revenue was computed as the number of saleable sandwiches (i.e., those with fewer than four mistakes) multiplied by the selling price. The selling price in the first period was fixed; the selling price in subsequent periods increased, decreased, or remained the same, depending on the average number of mistakes per sandwich in the previous period.⁹ Thus, the sandwich shop realizes long-term benefits from output quality.

Participants then read how their pay would be computed (described in the next section), and had to successfully complete a quiz to insure they understood the task. They then moved into the twelve four-minute production rounds. After the last work period, participants answered post-task questions to insure the task was not completed by an automated “bot”, and about their understanding of and beliefs about the task and their backgrounds.

Independent Variables

As with Study 1, we conducted two experiments in which payoff magnitude varied. In the *high pay* experiment, the first-period selling price of a sandwich was \$5, which is identical to that in Farrell et al. (2008). In the *low pay* experiment, the first-period selling price was \$1. Thus, possible earnings in the *high pay* experiment were equivalent to prior studies, while possible earnings in the *low pay* experiment were substantially smaller but more representative of pay levels in MTurk tasks.

We manipulated *contract type* at two levels between participants. For our *variable wage contract*, we use the *forward-looking contract* from Farrell et al. (2008). With the *variable wage*

⁹ Specifically, if the average number of mistakes per sandwich in a given period was zero, then the selling price in the next period was 10% higher; if the average number of mistakes per sandwich was greater than zero but less than two, then the selling price in the next period remained the same; and if the average number of mistakes per sandwich was two or more, then the selling price in the next period was 10% lower.

contract, participants were informed that they would be paid 5% of the revenue from the sale of perfect sandwiches (i.e., those with no errors) produced each period. We included a *flat wage contract* to assess baseline levels of effort and sensitivity to performance-based pay. These participants were informed that they would be paid \$25 in the *high pay* experiment and \$5 in the *low pay* experiment, regardless of production levels.¹⁰

Results

Attention, Comprehension, and Manipulation Checks

Recall that participants had to correctly answer all questions on a quiz before proceeding to the production task. In addition, in post-task questions, we asked participants to answer the question, “The number of mistakes that I made in one period a) affected my pay in subsequent periods, b) did not affect my pay in subsequent periods.” Across the *high* and *low pay* experiments, 74.5% of participants answered this question correctly. Results are the same if we exclude participants who did not answer this question correctly.¹¹

Dependent Variables

As in Farrell et al. (2008), we collect two measures of effort – production quality and quantity. Our proxy for quality is *average errors per sandwich*, computed as the mean across work periods of the total number of errors made each work period (excluding discarded sandwiches) divided by the number of saleable sandwiches in the period (lower values indicate higher quality). Our proxy for quantity is *average volume*, computed as the mean across all work periods of the number of saleable sandwiches produced (higher values indicate higher quantity).

¹⁰ Our experiment includes only the long-horizon conditions from Farrell et al. (2008), because including the short-horizon conditions would not significantly expand our understanding of our research questions.

¹¹ Most of the incorrect answers were in the *flat wage* condition. Because there is little reason why *flat wage* participants would believe that mistakes in one period affected their pay in subsequent periods, it is likely that incorrect answers reflected *flat wage* participants’ misunderstanding of the question as being about the *pricing* function rather than the *compensation* function.

We also calculate a measure of the total *revenue* earned by the principal (i.e., the sandwich shop). Because *revenue* is earned by the principal for all sandwiches with up to three mistakes—but participants only are rewarded for perfect sandwiches—this measure captures effort for which participants are only partially rewarded,

Hypothesis 2 Results – Effort Levels

Recall that H2 predicts that online workers will exhibit lower effort levels than found in prior research, regardless of whether their pay is comparable to that in prior research or not. We use the mean quality and quantity efforts from the Farrell et al. (2008) forward-looking variable wage contract as benchmarks (see Farrell et al. 2008, Table 1, Panel A, last column). Descriptive statistics are in Table 3.

----- Insert Table 3 about here -----

We first examine quality efforts. Note that lower *average errors* indicate higher quality effort. The benchmark from Farrell et al. (2008) is 0.23. In our *low pay* experiment, *average errors* were equivalent under the *flat wage contract* ($M_{\text{Farrell}}=0.10$ versus $M_{\text{OnlineFlatLow}}=0.10$, $t(45)=0.22$, $p=0.83$) and were lower under the *variable wage contract* ($M_{\text{Farrell}}=0.10$ versus $M_{\text{OnlineVariableLow}}=0.07$, $t(36)=-3.90$, $p<0.01$). Thus, even at significantly lower pay levels, online workers exhibited quality efforts that were at least comparable to, and sometimes higher than, those observed in prior research. As such, we find no support for H2. In our *high pay* experiment, *average errors* also are equivalent under the *flat wage contract* ($M_{\text{Farrell}}=0.10$ versus $M_{\text{OnlineFlatHigh}}=0.11$, $t(21)=0.19$, $p=0.85$) and lower under the *variable wage contract* ($M_{\text{Farrell}}=0.10$ versus $M_{\text{OnlineVariableHigh}}=0.06$, $t(21)=-2.80$, $p=0.01$). Thus, online workers' quality efforts were again at least equivalent to, if not greater than those in prior research with comparable wage levels. Importantly, this occurs even under a flat wage, rather than a

performance-based, contract. Again, results do not support H2.

We next examine quantity efforts. We use the *average volume* from Farrell et al. (2008) of 5.26 as our benchmark (see Farrell et al. 2008, Table 1, Panel C, third column). Across both *low* and *high pay* experiments, and across both the *variable wage and flat wage contracts*, *average volume* is significantly higher than our benchmark at the $p < 0.05$ level ($M_{\text{Farrell}}=5.51$ versus $M_{\text{OnlineFlatHigh}}=7.56$, $M_{\text{OnlineVariableHigh}}=7.60$, $M_{\text{OnlineFlatLow}}=6.12$, and $M_{\text{OnlineVariableLow}}=7.03$). Again, H2 is not supported.

In sum, we find that online workers exert comparable or higher levels of effort as those identified in prior research, regardless of the nature of the performance-based contract and even with a flat wage. It is especially notable that our participants could have earned comparable payoffs even if they had produced nothing. As such, we find no evidence that online workers are effort averse, and they appear to have sufficient motivations to work hard.

Hypothesis 3 Results – Pay-Performance Sensitivity

Recall that H3 predicts that online workers' efforts will be insensitive to differences in compensation contracts. We thus examine differences in quality and quantity efforts across contract types.

We tested whether online workers' efforts were consistent with purely economic predictions that effort will be higher under performance-based contracts than under a flat wage. Our results provide qualified support for these predictions. In our *low pay* experiment, relative to the flat wage contract, performance-based pay improved *average errors* ($M_{\text{OnlineVariableLow}}=0.07$ versus $M_{\text{OnlineFlatLow}}=0.10$, $t(81) = -2.06$, $p=0.04$), *average volume* ($M_{\text{OnlineVariableLow}}=7.03$ versus $M_{\text{OnlineFlatLow}}=6.12$, $t(81) = 2.26$, $p=0.03$), and *revenue* ($M_{\text{OnlineVariableLow}}=\10.97 versus $M_{\text{OnlineFlatLow}}=\9.26 , $t(81) = 2.60$, $p=0.01$). By contrast, in our *high pay* experiment, we observe

no differences in any of the three measures across the performance-based and the flat wage contracts (all $p > 0.31$). In combination with our results from H2, these results suggest that performance-based pay is not necessary to induce online workers to exert reasonable effort under certain conditions.

Thus, our results provide qualified support for H3. Online workers in our study are only sensitive to incentive contract differences when payoffs are relatively lower. They are not at all sensitive to contract type when payoffs are relatively higher. We suspect that this could be due either to online workers' incentives to perform well to protect their work reputations when payoffs are high or to incremental enjoyment of the task that results from relatively pay high.

Taken together, results for H2 and H3 indicate that workers in online labor markets exert a considerable degree of effort. Even when pay is much lower than that offered in traditional laboratory environments, and even with a flat wage, efforts are at least comparable to those found in prior research and similarly respond to differences in incentive contracts. When pay is roughly equivalent to payoffs in laboratory environments, efforts are also comparable to prior research, but, interestingly, insensitive to contract type. This is consistent with the idea that performance-contingent rewards are not likely to be as effective when tasks are inherently interesting to online workers, and that reputation effects become more important when online workers want to repeatedly engage with a high-paying employer.

However, in study 2 we cannot disentangle whether our findings for online workers' efforts are due to the inherently interesting nature of the task used, or the salience of reputation concerns in the online market. We more closely examine these issues of online workers' motivations in study 3.

VI. STUDY 3 – EFFORT AND PAY-PERFORMANCE SENSITIVITY ON A TASK THAT IS NOT GAME-LIKE

Study 3 involves a task that is not game-like, and has two objectives. The first objective is to further examine online workers' sensitivity to variable wages. If online workers' relative insensitivity to variable wages is driven by their stronger intrinsic motivation on game-like tasks, then their sensitivity should be more apparent in a task that is not game-like. The second objective is to more precisely examine the differences in motivations between online workers and students. To do so, we include groups of online workers and groups of students in study three. This enables us to examine (1) whether online workers maintain their performance advantage over students on a task that is not game-like, when online workers' intrinsic motivation advantage is likely diminished, and (2) whether performance differences between online workers and students are the result of differences in intrinsic enjoyment of games, concerns about reputation, or both.

Procedures

As in study 2, this study was designed using custom software on a private web site. Instructions and procedures were nearly identical to those in study 2, and we note exceptions below. Participants completed orders using a computerized order and assembly system for four minute periods. They received detailed instructions and completed a practice round. The assembly screen was also the same as in study 2, with five drop down menus and 51 unique combinations of "ingredients." The pricing and revenue calculations were identical to study 2, as was the rule that anything with four or more mistakes would be discarded. The dependent variables were identical to those in study 2, as were the *low pay* and *high pay* levels.

In contrast to study 2, instead of sandwiches we asked participants to assume that they were completing product orders that consisted of parts represented by shapes, i.e., hexagons, squares, circles, pentagons, and triangles, instead of sandwich ingredients. Instead of sandwich names, each product order was simply given a number between 1 and 51. In the drop down menus, each part was labeled with a shape and number, e.g., “hexagon 5” or “square 6.” As participants assembled the order by adding parts, the shape and its label would appear on the assembly screen. In contrast to the sandwich ingredients in study 2, the product parts were not colorful and did not fit together in any organized fashion. The parts appeared in the window in the order in which they were added. This design was intended to minimize the “game-like” feeling of the task, and to maximize participants’ feeling that this was a straightforward production task.

Dependent Measures

Our dependent measures of error, volume, efficiency, and revenue are identical to those in study 2. In this study, we also collected measures of participants’ motivations during the task. On a scale of 0 (“strongly disagree”) to 10 (“strongly agree”), they assessed their agreement with the statement “I enjoy playing computer games.” Also, on a scale of 0 (“not at all concerned”) to 10 (“extremely concerned”) participants assessed the degree to which they were concerned with their reputation with the requester (seen by online workers) / researcher (seen by students).

Independent Variables

As in study 2, we conducted two experiments in which we manipulated the payoff magnitude. The *low pay* and *high pay* levels were identical to those in study 2. We also manipulated contract type at two levels between participants. In the *flat wage* condition, participants earned \$25 in the *high pay* experiment and \$5 in the *low pay* experiment, regardless

of production levels. In the *variable wage* condition, *participants* were told that they would earn 5% of the revenue from the sale of perfect products (i.e., those with no errors) produced in that period.

Our third independent variable is whether the participant was recruited from MTurk or from the population of upper-level undergraduate business students at a large university in the Midwest U.S. Online workers were recruited using identical procedures to those used in study 2. We selected students from several sections of business courses and sent an email solicitation for their participation after the end of the semester.

Results

As in the game-like task in study 2, online workers appear to exhibit limited sensitivity to variable wages in study 3. See Table 3 for cell means. Online workers' error is lower in the variable wage condition than in the flat wage condition for the high pay contract ($M_{\text{OnlineVariableHigh}} = 0.05$ versus $M_{\text{OnlineFlatHigh}} = 0.08$, $t(41) = -2.01$, $p = 0.06$) and marginally lower for the low pay contract ($M_{\text{OnlineVariableLow}} = 0.06$ versus $M_{\text{OnlineFlatLow}} = 0.09$, $t(78) = -1.64$, $p = 0.10$). Further, variable wages marginally increase the principal's revenue when pay is low ($M_{\text{OnlineVariableLow}} = \10.78 versus $M_{\text{OnlineFlatLow}} = \9.57 , $t(78) = 1.70$, $p = 0.09$). In sum, these results support the idea that online workers' effort exhibits qualified sensitivity to variable wages, and we observed this effect on a task that is game-like and a task that is not game-like.

In direct comparisons between online workers and students, the effort of the two groups was equivalent on all three measures and under both pay levels. This is consistent with the findings in study 2 that online workers are no worse than students on any of our measures. It is possible that online workers' performance advantage diminished because the task was less game-like and, thus, their intrinsic motivation advantage over the students diminished, as well.

We now turn to an examination of the underlying differences in the motivations of online workers and students. As discussed in Section 2, online workers are likely to have greater intrinsic motivation and are also likely to be more strongly motivated by reputation concerns, relative to students. Indeed, online workers report greater enjoyment of computer games ($M_{\text{Online}}=8.61$ versus $M_{\text{Student}}=5.94$, $t(169)=6.78$, $p < 0.01$) and stronger reputation concerns ($M_{\text{Online}}=5.92$ versus $M_{\text{Student}}=3.71$, $t(169)=4.06$, $p < 0.01$). Either of these factors could lead to differences in the efforts of online workers, as opposed to student participants. That is, online workers' effort may differ from that of students because online workers enjoy computer tasks more, because online workers are more concerned about their reputations, or both. To disentangle the effects of these variables, we examine whether enjoyment and reputation concerns mediate the relation between participant type and effort. To do so, we conduct three indirect effects analyses—one for each of the three dependent measures—in which participant type is the independent variable and both reputation concerns and enjoyment are included as mediators.

Following the recent statistics literature on mediation, we test the significance of the indirect effects using a bootstrapping technique developed by Preacher and Hayes (2008).¹³ In each of the four analyses, we use 5000 bootstrap re-samples of the data to calculate bias-corrected 95% confidence intervals for the total indirect effect via reputation concerns and via computer game enjoyment. The indirect effect is the product of paths *a* and *b* in Figure 1. Significance is indicated by confidence intervals that do not include zero. The analyses support the conclusion that enjoyment explains the effort differences between online workers and students in our sample, as enjoyment significantly mediates this relation for two of the three

¹³ This technique has several advantages over the traditional Baron and Kenny stepwise approach, including that it is robust to violations of normality and reduces the number of inferential tests, thus minimizing the likelihood of Type I error. Our inferences are identical if we use the Baron and Kenny approach.

measures. That is, online workers' greater enjoyment relative to students increases online workers' volume (0.23, 1.01) and revenue (1.31, 8.98).

By contrast, reputation concerns mediate the relation between participant type and performance only on the effort measure. Surprisingly, to the extent that online workers' stronger reputation concerns explain differences in performance, these stronger reputation concerns *worsen* online workers' performance. See Figure 1. This indirect effect of reputation, via reputation concerns, is positive (0.01, 0.02). That is, being an online worker increases reputation concerns and this, in turn, increases error. In sum, reputation concerns worsen performance on the only measure for which they affect performance, while enjoyment improves performance on two of the three measures. Thus, our data support the conclusion that the performance differences between online workers and students are driven by online workers' greater enjoyment of the task.

VII. DISCUSSION AND CONCLUSIONS

Online labor markets offer purchasers of labor quick access to diverse labor pools in which workers are often willing to work for relatively small sums of money. However, we know little about contracting issues within these markets, particularly in regards to the underlying quality of labor. We examine whether online workers report honestly in settings with high information asymmetries; whether they are willing to work hard on tasks that are relatively complex and time-consuming; and whether their efforts are sensitive to performance-based pay, a common contracting consideration.

In three studies, we find that online workers exhibit preferences for honesty over wealth that are comparable to those found in prior research, regardless of the level of economic incentives to lie. We also find that online workers exert effort to the same, if not a greater, extent than found in

prior research, even when online workers are paid a flat wage. Further, online workers' efforts become less sensitive to differences in compensation contracts as pay levels increase. Finally, we find that online workers have higher intrinsic motivations than do traditional lab participants. In turn, this motivational difference improves online workers' performance.

Overall, our evidence suggests that the quality of labor in online labor markets is reasonably high, and potentially higher than theory would predict given the potential for adverse selection that is present in these markets. For instance, in our experiments across three different tasks and across several measures, the quality of online workers was never worse than that of student participants. Online workers' exhibited superior quality in some instances. While it does not necessarily prove that online workers are stars because they equaled the performance of college students, it is noteworthy that online workers did so for substantially lower wages. Thus, our findings raise the possibility that eliciting higher effort from online workers may cost less than doing so with other populations, at least under some conditions. In brief, the costs of aligning online workers' motivations with the principal's interest are likely to be lower than one may think.

It is important to note that we examined contracting with online workers on MTurk, where adverse selection problems would theoretically be worse than most other online markets—e.g., workers are completely anonymous and wages are very low. If labor quality is higher than expected on MTurk, then it is reasonable to generalize this inference to other online labor markets, as well.

Another notable finding in our study is that online workers appear to be relatively sensitive to performance-based wages at lower pay levels but not at higher pay levels. Among some online workers, economic incentives may not be necessary to motivate high performance. To the extent

that incentives do work, contract designers should note that the benefits of performance-based pay are more likely to be realized when wage levels are relatively low. However, our findings do not suggest that it is optimal to simply set pay levels very low when contracting with online works. While the lower pay levels in our paper—\$5 to \$6 for an hour-long task—are low relative to many real-world and laboratory contracting settings, they are substantially higher than pay levels in many other online contracting settings.

Also, our results offer multiple pieces of guidance to accounting researchers considering the relative benefits and costs of using online workers as participant pools for long, complex accounting tasks. First, a stream of studies in accounting examines preferences for honesty over wealth (Evans et al. 2001; Hannan, Rankin and Towry 2006; Zhang 2008; Church, Hannan and Kuang 2012; Maas and van Rinsum 2013). Our results indicate that researchers interesting in accounting domains in which honesty preferences are an important or assumed construct – for example, tax evasion, financial reporting choice, and auditing, where dishonesty can have severe consequences for firms, markets, and governments – can comfortably consider using workers in online labor markets in their studies.

Second, accounting researchers can also comfortably consider using online labor markets for settings in which high effort expenditures or high intrinsic motivation are required. Our results indicate that comparable or greater effort can be induced at substantially lower wages than those paid in prior research, although certainly fair wages should be paid for the work requested. As such, we contribute to discussions about the suitability of particular participant pools for various research populations of interest by identifying a potentially cost-effective, quick source for reliable data (e.g., Peecher and Solomon 2001; Libby, Bloomfield and Nelson 2002; Elliott, Hodge, Kennedy and Pronk 2007; Henrich, Heine and Norenzayan 2010).

Third, we offer a unique perspective on the effects of the nature of the task on performance in accounting research tasks. Many accounting studies have moved toward the use of interesting and engaging tasks to test the effects of economic incentives on performance (e.g., Farrell et al. 2008, 2012; Kachelmeier et al. 2008; Kachelmeier and Williamson 2010; Chen, Williamson and Zhou 2012; Choi et al. 2012, 2013). However, our results suggest that in some populations, it is possible that these tasks intrinsically engage research participants to such an extent that economic incentives may not be necessary to motivate high performance.

Our study has limitations that represent opportunities for future research. First, we selected only a few important constructs used in contracting research, so our investigation of the characteristics and decisions of online workers is by no means exhaustive. For example, contracting in team settings differs from contracting in individual settings. It possible that online teams perform less effectively than laboratory teams due to the lack of team affiliation and shared goals, but it is also possible that crowdsourcing leads to more effective performance on problem-solving tasks because it eliminates the interpersonal challenges of teamwork. Future research could examine this issue. Second, we limited participation to U.S. participants; it is possible that differences across nationality, culture, or language may impact our results.

REFERENCES

- Amazon.com. 2014a. Amazon Mechanical Turk Requester Service Summary. Web. 26 March 2014 <<https://requester.mturk.com/tour>>.
- Amazon.com. 2014b. FAQ: Worker Web Site FAQs. Web. 25 March 2014 <<https://www.mturk.com/mturk/help?helpPage=worker>>.
- Baiman, S. and J. Evans. 1983. Pre-decision information and participative management control systems. *Journal of Accounting Research* 21 (2): 371–395.
- Bazerman, M., J. Curhan, D. Moore, and K. Valley. 2000. Negotiation. *Annual Review of Psychology* 51: 279-314.
- Bohnet, I. and B. Frey. 1999. The sound of silence in prisoner's dilemma and dictator games. *Journal of Economic Behavior and Organizations* 38: 43-57.
- Brazel, J., C. Agoglia, and R. Hatfield. 2004. Electronic versus face-to-face review: The effects of alternative forms of review on auditors' performance. *The Accounting Review* 79(4): 949-966.
- Brink, W. 2013. The effects of sharing the consequences of tax evasion and regret salience on taxpayer compliance. Working paper – Miami University.
- Brink, W. and L. Lee. 2013. The effect of tax preparation software on aggressive reporting. Working paper – Miami University.
- Buhrmester, M., T. Kwang and S. Gosling. 2011. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science* 6(1): 3-5.
- Chandler, J., P. Mueller, and G. Paolacci. 2014. Nonnaivete among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods* 46(1): 112-130.
- Chen, E. 2012. Making the most of Mechanical Turk: Tips and best practices. Web. 25 March 2014 <<http://blog.echen.me/2012/04/25/making-the-most-of-mechanical-turk-tips-and-best-practices/>>.
- Chen, X., and T. Sandino. 2012. Can wages buy honesty? The relationship between relative wages and employee theft. *Journal of Accounting Research* 50(4): 967-1000.
- Chen, X., M. Williamson, and F. Zhou. 2012. Reward system design and group creativity: An experimental investigation. *The Accounting Review* 87(6): 1885-1911.
- Choi, J., G. Hecht, and W. Tayler. 2012. Lost in translation: The effects of incentive compensation on strategy surrogation. *The Accounting Review* 87(4): 1135-1163.
- Choi, J., G. Hecht, and W. Tayler. 2013. Strategy selection, surrogation, and strategic performance measurement systems. *Journal of Accounting Research* 51(1): 105-133.
- Choi, J., A. Newman, and I. Tafkov. 2013. A marathon, a series of sprints, or both? Tournament horizon and task complexity in multi-period settings. Working paper – University of Pittsburgh, University of South Carolina, and Georgia State University.
- Church, B., R. Hannan, and X. Kuang. 2012. Shared interest and honesty in managerial reporting. *Accounting, Organizations and Society* 37(3): 155-167.

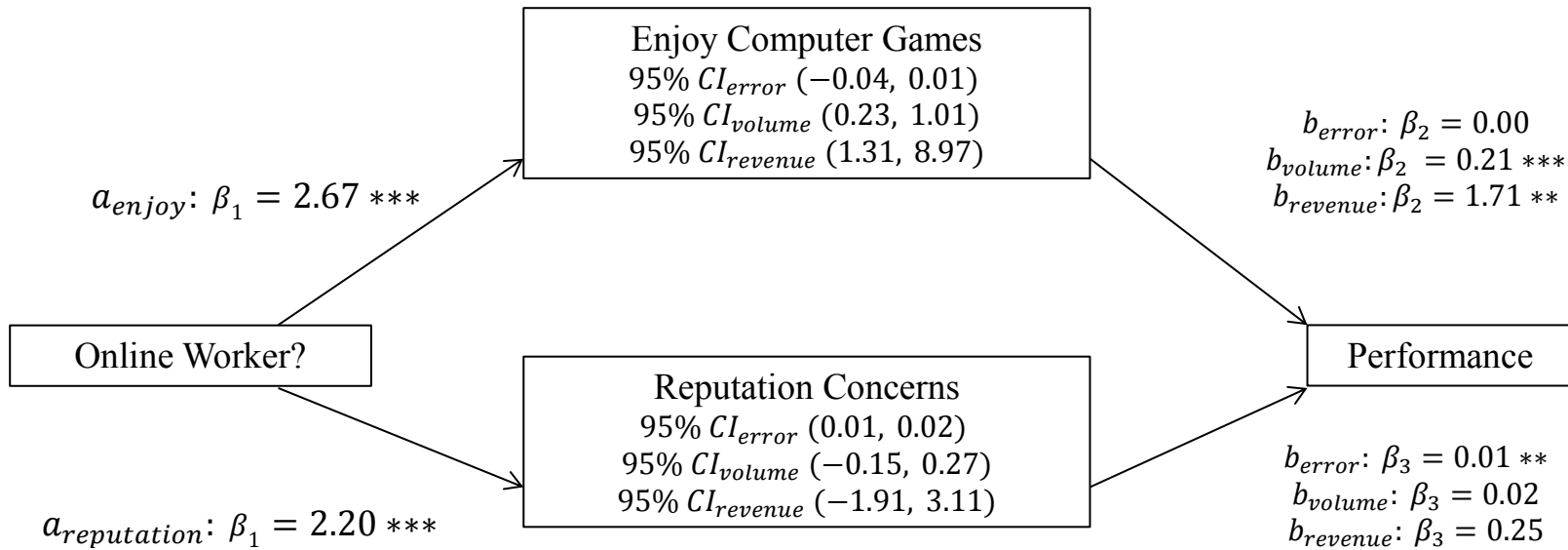
- Crump, M., J. McDonnell, and T. Gureckis. 2013. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLOS ONE* 8(3): 1-18.
- Deci, E., R. Koestner, and R. Ryan. 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* 125(6): 627-668.
- Dickinson, D. and M. Villeval. 2008. Does monitoring decrease work effort? The complementarity between agency and crowding-out theories. *Games and Economic Behavior* 63: 56-76.
- Dobson, J. 2013. Mechanical Turk: Amazon's new underclass. *The Huffington Post* 19 February. Web. 25 March 2014. <http://www.huffingtonpost.com/julian-dobson/mechanical-turk-amazons-underclass_b_2687431.html>.
- Elliott, W., F. Hodge, J. Kennedy, and M. Pronk. 2007. Are M.B.A. students a good proxy for nonprofessional investors? *The Accounting Review* 82(1): 139-168.
- Evans, J., R. Hannan, R. Krishnan, and D. Moser. 2001. Honesty in managerial reporting. *The Accounting Review* 76(4): 537-559.
- Farrell, A., K. Kadous, and K. Towry. 2008. Contracting on contemporaneous versus forward-looking: An experimental investigation. *Contemporary Accounting Research* 25(3): 773-802.
- Farrell, A., K. Kadous, and K. Towry. 2012. Does the communication of causal linkages improve effort allocations and firm performance? An experimental investigation. *Journal of Management Accounting Research* 24: 77-102
- Fehr, E. and A. Falk. 2002. Psychological foundations of incentives. *European Economic Review* 46(4-5): 687-724.
- Fehr, E. and U. Fischbacher. 2002. Why social preferences matter – The impact of non-selfish motives on competition, cooperation, and incentives. *The Economic Journal* 112(478): C1-C33.
- Fehr, E., and K. Schmidt. 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114 (3): 817–868.
- Folbre, N. 2013. The unregulated work of Mechanical Turk. *The New York Times* 18 March. Web. 25 March 2014. <http://economix.blogs.nytimes.com/2013/03/18/the-unregulated-work-of-mechanical-turk/?_php=true&_type=blogs&_r=0>.
- Glazer, E. 2011. Serfing the web: Sites let people farm out their chores. *The Wall Street Journal* 28 November. Web. 26 March 2014. <<http://online.wsj.com/news/articles/SB10001424052970204443404577052353225234154>>
- Gneezy, U., and A. Rustichini. 2000. Pay enough or don't pay at all. *The Quarterly Journal of Economics* 115(3): 791-810.
- Goodman, J., C. Cryder, and A. Cheema. 2012. Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*
- Grenier, J., D. Lowe, A. Reffett, and R. Warne. 2014. The effects of expert panel recommendations on juror judgments of auditor negligence. Working Paper – Arizona State University, George Mason University, and Miami University.

- Grenier, J., B. Pomeroy, and M. Stern. 2014. The effects of accounting standard precision, auditor task expertise, and judgment frameworks on audit firm litigation exposure. *Forthcoming - Contemporary Accounting Research*.
- Hannan, R., F. Rankin, and K. Towry. 2006. The effect of information systems on honesty in managerial reporting: A behavioral perspective. *Contemporary Accounting Research* 23(4): 885-918.
- Henrich, J., S. Heine, and A. Norenzayan. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33: 61-135.
- Hoffman, E., K. McCabe, and V. Smith. 1996. Social distance and other-regarding behavior in dictator games. *American Economic Review* 86(3): 653-660.
- Holmes, E. 2009. Selling expertise on the internet for cash. *The Wall Street Journal* 3 March. Web. 26 March 2014. <<http://online.wsj.com/news/articles/SB123604090494214861>>.
- Horton, J. and L. Chilton. 2010. The labor economics of paid crowdsourcing. *Proceedings of the 11th ACM Conference on Electronic Commerce*. New York, NY: ACM.
- Horton, J., D. Rand, and R. Zeckhauser. 2011. The online laboratory: Conducting experiments in real labor markets. *Experimental Economics* 14: 399-425.
- Kachelmeier, S., B. Reichert, and M. Williamson. 2008. Measuring and motivating quantity, creativity, or both. *Journal of Accounting Research* 46(2): 341-373.
- Kachelmeier, S. and M. Williamson. 2010. Attracting creativity: The initial and aggregate effects of contract selection on creativity-weighted productivity. *The Accounting Review* 85(5): 1669-1691.
- Kachelmeier, S. and K. Towry. 2002. Negotiated transfer pricing: Is fairness easier said than done? *The Accounting Review* 77(3): 571-593.
- Koonce, L., J. Miller, and J. Winchel. 2013. The effects of norms on investor reactions to derivative use. Working paper – University of Notre Dame, University of Texas, and University of Virginia.
- Libby, R., R. Bloomfield, and M. Nelson. 2002. Experimental research in financial accounting. *Accounting, Organizations and Society* 27(8): 775-810.
- Maas, V., and M. van Rinsum. 2013. How control system design influences performance misreporting. *Journal of Accounting Research* 51(5): 1159-1186.
- Mason, W., and S. Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavioral Research Methods* 44(1): 1-23.
- McGonigal, J. 2011. *Reality is Broken: Why Games Make Us Better and How They Can Change the World*. New York, NY: Penguin Books.
- Melumad, N. and S. Reichelstein. 1987. Centralization vs. delegation and the value of communication. *Journal of Accounting Research* 25 (Supplement): 1-18.
- MTurk Tracker. 2014. General Data: Hits, January 1-December 31, 2013. Web. 26 March 2014 <http://mturk-tracker.com/general/?date_from=01%2F01%2F2013&date_to=12%2F31%2F2013>.

- Oppenheimer, D., T. Meyvis, and N. Davidenko. 2009. Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology* 46(4): 867-872.
- Paolacci, G., J. Chandler, and P. Ipeirotis. 2010. Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5(5): 411-419.
- Pattison, K. 2009. Enlisting a global work force of freelancers. *The New York Times* 24 June. Web. 26 March 2014.
<http://www.nytimes.com/2009/06/25/business/smallbusiness/25freelance.html?_r=1>.
- Peecher, M., A. Reffett, and A. Zimbelman. 2014. Juror dysfunction: Can auditors reduce the penalty for narrowly failing to detect fraud? Working Paper – Miami University and University of Illinois at Urbana-Champaign.
- Peecher, M. and I. Solomon. 2001. Theory and experimentation in studies of audit judgments and decisions: Avoiding common research traps. *International Journal of Auditing* 5: 193-203.
- Penno, M. 1984. Asymmetry of pre-decision information and managerial accounting. *Journal of Accounting Research* 22 (1): 177–191.
- Pontin, J. 2007. Artificial intelligence, with help from humans. *The New York Times* 25 March. Web. 26 March 2014.
<<http://www.nytimes.com/2007/03/25/business/yourmoney/25Stream.html>>.
- Rankin, F., S. Schwartz, and R. Young. 2008. The effect of honesty and superior authority on budget proposals. *The Accounting Review* 83(4): 1083-1099.
- Rennekamp, K. 2012. Processing fluency and investor reactions’ to disclosure readability. *Journal of Accounting Research* 50(5): 1319-1354.
- Rennekamp, K., K. Rugar, and N. Seybert. 2013. Impaired judgment: The effects of asset impairment reversibility and cognitive dissonance on future investment. *The Accounting Review*, forthcoming.
- Roggio, A. 2012. Try Amazon’s Mechanical Turk for inexpensive labor. Web. 25 March 2014
<<http://www.practicalcommerce.com/articles/3282-Try-Amazon-s-Mechanical-Turk-for-Inexpensive-Labor>>.
- Ross, J., L. Irani, M. Silberman, A. Zaldivar, and B. Tomlinson. 2010. Who are the crowdworkers? Shifting demographics in Amazon Mechanical Turk. *CHI ‘10 Extended Abstracts on Human Factors in Computing Systems*, pages 2863-2872. New York, NY: ACM.
- Ryan, R., and E. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55(1): 68-78.
- Silverman, R. 2012. Big firms try crowdsourcing. *The Wall Street Journal* 17 January. Web. 26 March 2014.
<<http://online.wsj.com/news/articles/SB10001424052970204409004577157493201863200>>
- Sprinkle, G. 2003. Perspectives on experimental research in managerial accounting. *Accounting, Organizations and Society* 28(2-3): 287-318.

- Suri, S., D. Goldstein, and W. Mason. 2011. Honesty in an online labor market. *Human Computation Papers from the 2011 AAAI Workshop*. Palo Alto, CA: AAAI.
- Valley, K., J. Moag, and M. Bazerman. 1998. 'A matter of trust': Effects of communication on the efficiency and distribution of outcomes. *Journal of Economic Behavior and Organization* 34: 211-238.
- Worstell, T. 2013. On the New York Times' stupidity over Amazon's Mechanical Turk. *Forbes* 19 March. Web. 26 March 2014. <<http://www.forbes.com/sites/timworstell/2013/03/19/on-the-new-york-times-stupidity-over-amazons-mechanical-turk/>>.
- Young, S. and B. Lewis. 1995. Experimental incentive-contracting research in management accounting. In R. Ashton and A. Ashton (Eds.), *Judgment and Decision Making Research in Accounting and Auditing*. Cambridge, UK: Cambridge University Press.
- Zhang, Y. 2008. Effects of fairness and communication on honesty and collusion in a multi-agent setting. *The Accounting Review* 83(4): 1125-1146.

**FIGURE 1
Mediation**



*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

This figure depicts the coefficients of the indirect effect of participant type on performance, via enjoyment and reputation concerns. The coefficients for the paths denoted a are from this regression, conducted separately for enjoyment and reputation concerns:

$$Mediator = \alpha + \beta_1 OnlineWorker? + \varepsilon$$

The coefficients for the paths denoted b are from this regression, conducted separately for each performance measure:

$$Performance = \alpha + \beta_2 Enjoyment + \beta_3 Reputation + \beta_4 Online Worker? + \varepsilon$$

β_1 , β_2 , and β_3 are reported in the figure. The *indirect effect* of the independent variable on performance is $\beta_1 * \beta_2$ for the enjoyment mediator and $\beta_1 * \beta_3$ for the reputation mediator. β_4 is the *direct effect* of the independent variable is not reported as it is not relevant to this analysis. Online worker? = 0 for student participants and 1 for MTurk participants. Confidence intervals are bias-corrected intervals for the estimate of the indirect effect, which are estimated using 5,000 bootstrapped re-samples of the data with replacement. Significance of the indirect effect is indicated if the intervals do not include zero.

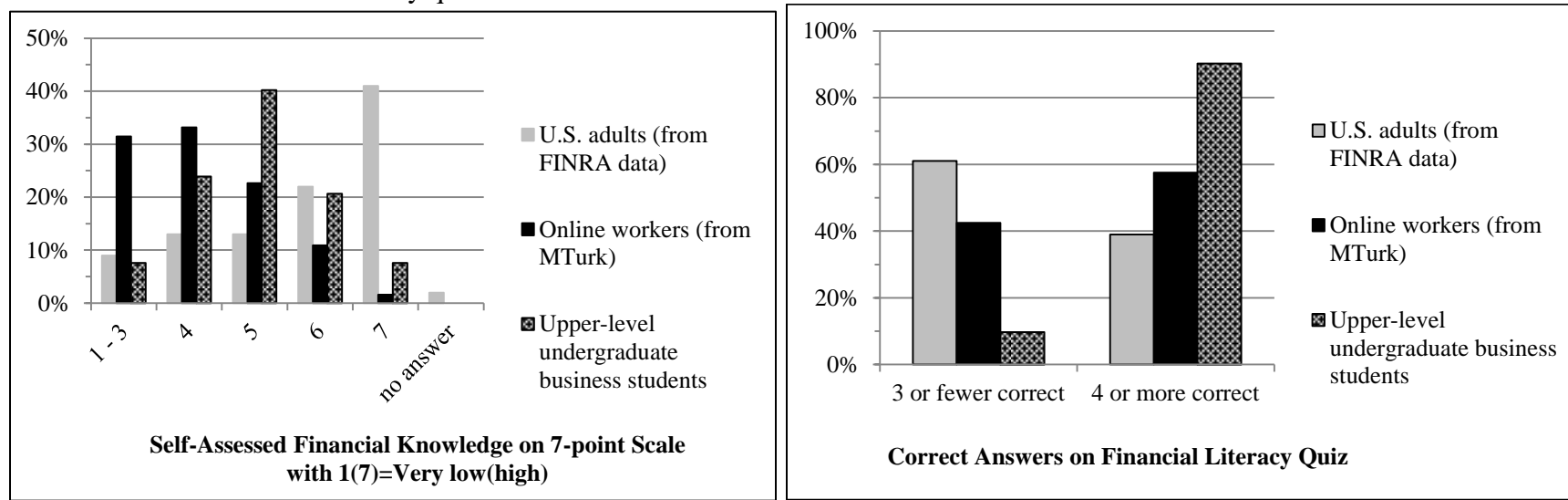
TABLE 1
Demographic Characteristics of Online Workers from MTurk

Panel A: Demographic characteristics of Online Workers from MTurk in prior studies and Experiments 1 and 2

	Paolacci et al. (2010)	Ross et al. (2010), Nov. 2009 study	Buhrmester et al. (2011)	Goodman et al. (2012), Study 1 / 2	Experiment 1	Experiment 2	Experiment 3
Sample size	1,000	733	3,006	107 / 207	128	127	122
Proportion U.S.-based	47%	56%	69%	74% / 37%	100%	100%	100%
Average age	36 years for U.S.-based	35 years for U.S.-based	33 years	33 / 31 years	---	---	---
Household income	for U.S.-based: <ul style="list-style-type: none"> • 67% < \$60,000/year • distribution matches U.S. population 	for U.S.-based: <ul style="list-style-type: none"> • 10% < \$10,000 • 24% \$10,000-\$30,000 • 25% \$30,000-\$50,000 • 18% \$50,000-\$70,000 • 23% > \$70,000 	---	---	---	---	---
Percent female	for U.S.-based, 65%	for U.S.-based, 63%	55%	59% / 43%	44%	53%	43%
English as first language	---	---	---	---	92%	95%	98%
Education	for U.S.-based, higher than that of U.S. population	38% Bachelor's, 17% graduate	---	Median: Bachelor's / Bachelor's	<ul style="list-style-type: none"> • 38% Bachelor's • 4% Graduate • of those with degree, 39% in business or economics 	<ul style="list-style-type: none"> • 35% Bachelor's • 8% Graduate • of those with degree, 15% in business or economics 	<ul style="list-style-type: none"> • 36% Bachelor's • 12% Graduate • Of those with degree, 18% in business or economics
Have managerial experience	---	---	---	---	37%	39%	34%
Employment status	---	---	---	---	---	<ul style="list-style-type: none"> • 42% full-time • 13% part-time • 23% self-employed 	<ul style="list-style-type: none"> • 45% full-time • 10% part-time • 24% self-employed

TABLE 1 (continued)

Panel B: FINRA financial literacy quiz ^a



Notes:

- a The financial literacy quiz was administered by the FINRA Investor Education Foundation as part of a 2012 study of over 25,000 American adults. We included the quiz in post-task questions for Experiments 1, 2, and 3, and separately administered the quiz in paper-and-pencil form to 42 upper-level undergraduate business students in an advanced auditing class. The first graph presents responses to the question, “How would you assess your overall financial knowledge?”, using a seven-point scale anchored on 1(7)=Very low (high). The second graph presents the percentage of participants who correctly answered three of fewer versus four or more of the following five questions (correct answers in bold):
- Suppose you have \$100 in a savings account earning 2% interest per year. After five years, how much would you think you would have in the account if you left the money to grow? (Response choices: **More than \$102**, Exactly \$102, Less than \$102, Don’t know)
 - Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After one year, how much would you be able to buy with the money in this account? (Response choices: More than today, Exactly the same, **Less than today**, Don’t know)
 - If interest rates rise, what will typically happen to bond prices? (Response choices: They will rise, **They will fall**, They will stay the same, There is no relationship between bond prices and the interest rate, Don’t know)
 - True or false: A 15-year mortgage typically requires higher monthly payments than a 30-year mortgage but the total interest over the life of the loan will be less. (Response choices: **True**, False, Don’t know)
 - True or false: Buying a single company’s stock usually provides a safer return than a stock mutual fund. (Response choices: True, **False**, Don’t know)

TABLE 2
Hypothesis 1 Results – Honesty

	Online Workers		Evans et al. (2001), Experiments 1 and 3
	Low Pay (~1/5 of Evans et al. 2001)	High Pay (comparable to Evans et al. 2001)	
<i>Trust Contract</i>			
N	470	230	280
Percentage of Honesty ^a	49.9% ^{c, d}	57.3% ^{c, d}	48.7% [benchmark]
Payoff foregone ^b			
Mean	\$3.39	\$19.47	\$16.07
Maximum	\$6.80	\$34.00	\$61.67
Report Types			
Fully Dishonest	15%	0%	27%
Fully Honest	4%	4%	25%
<i>Modified Trust Contract</i>			
N	390	190	142
Percentage of Honesty ^a	24.0% ^{c, d}	15.3% ^{c, d}	21.8% [benchmark]
Payoff foregone ^b			
Mean	\$0.41	\$1.29	\$3.61
Maximum	\$1.70	\$8.00	\$31.67
Report Types			
Fully Dishonest	49%	53%	61%
Fully Honest	5%	0%	16%

Notes:

- a The *percentage of honesty* is computed as: $1 - (\sum_{i=1}^n \text{Payoff Claimed} / \sum_{i=1}^n \text{Payoff Available})$.
- b The *payoff foregone* is the extra wealth a participant *could* have claimed by reporting fully dishonestly. It is equal to the sum of the available payoffs less the payoffs actually claimed.
- c Across both the *low* and *high pay* experiments, and within both the *trust* and *modified trust contracts*, *percentage of honesty* does not statistically differ from the comparable benchmark percentage reported in Evans et al. (2001).
- d Within both the *low* and *high pay* experiments, *percentage of honesty* is significantly higher with the *trust contract* than the *modified trust contract*.

TABLE 3
Hypotheses 2 and 3 – Effort and Pay-Performance Sensitivity

Panel A: Experiment 2, Game-Like Task (Sandwich Shop)

	<u>High Pay</u>			<u>Low Pay</u>	
	Farrell et al. 2008 [benchmark] ^d	Online Workers			
		Flat Wage ^e (not in Farrell et al. 2008)	Variable Wage	Flat Wage (not in Farrell et al. 2008)	Variable Wage
N	20	22	22	46	37
Mean (std. dev.) of:					
Average errors ^a	0.10 (0.09)	0.11 (0.19)	0.06 (0.05)	0.10† (0.09)	0.07† (0.12)
Average volume ^b	5.26 (1.02)	7.56 (1.33)	7.60 (1.53)	6.12† (2.03)	7.03† (1.78)
Revenue ^c	\$38.59 (8.92)	\$56.99 (14.04)	\$59.90 (16.16)	\$9.26† (3.37)	\$10.97† (3.16)
Average compensation	\$25.55	\$25.00	\$34.81	\$5.00	\$6.32

Panel B: Experiment 3, Non Game-Like Task (Product Assembly)

	<u>High Pay</u>				<u>Low Pay</u>	
	Students ^d		Online Workers			
	Flat Wage ^e	Piece Rate	Flat Wage	Piece Rate	Flat Wage	Piece Rate
N	26	23	24	19	36	44
Mean (std. dev.) of:						
Average errors ^a	0.13 (0.22)	0.08 (0.07)	0.08† (0.07)	0.05† (0.05)	0.09† (0.09)	0.06† (0.07)
Average volume ^b	6.50 (1.39)	6.88 (1.59)	7.24 (1.90)	7.22 (2.33)	6.32 (2.25)	6.85 (1.90)
Principal's revenue ^c	\$49.80 (13.08)	\$55.34 (13.68)	\$55.81 (15.60)	\$59.70 (18.28)	\$9.57† (3.10)	\$10.78† (3.29)
Average compensation	\$25.00	\$31.84	\$25.00	\$34.92	\$5.00	\$6.26

Notes:

- a Average errors is computed as the mean across work periods of the total number of errors made each work period (excluding discarded sandwiches), divided by the number of saleable sandwiches in the period. Lower values indicate higher quality efforts.
- b Average volume is computed as the mean across all work period of the number of saleable sandwiches produced. Higher values indicate higher quantity efforts.
- c Revenue is the average revenue per period that the principal earns from participants' production. Higher values indicate higher revenue.
- d For online workers in Panel A, ***boldface italic*** cell means indicate that online workers' mean is statistically different from the benchmark mean from prior research. For online workers and students in Panel B, ***boldface italic*** cell means indicate that the online workers' mean is statistically different from students' cell mean for the given contract type.
- e For the flat wage and piece rate conditions, a † next to the cell means indicate that the flat wage mean is significantly different from the piece rate cell mean for the given participants and pay level.