

IZA DP No. 6410

**The Multi-Dimensional Effects of Reciprocity
on Worker Effort: Evidence from a Hybrid
Field-Laboratory Labor Market Experiment**

Min-Taec Kim
Robert Slonim

March 2012

The Multi-Dimensional Effects of Reciprocity on Worker Effort: Evidence from a Hybrid Field-Laboratory Labor Market Experiment

Min-Taec Kim

University of Sydney

Robert Slonim

*University of Sydney
and IZA*

Discussion Paper No. 6410

March 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

The Multi-Dimensional Effects of Reciprocity on Worker Effort: Evidence from a Hybrid Field-Laboratory Labor Market Experiment*

We examine the gift exchange hypothesis on both the quantity and *quality* of output using a hybrid field-laboratory labor market experiment. We recruited participants to enter survey data for a well-known charitable organization. Workers were paid either a *high* or *low* wage. We find that although the total number of surveys entered did not vary with the wage, high wage workers made fewer errors and entered more surveys after controlling for errors. We further find that for low costs associated with errors, offering the low wage maximizes profits, but for higher costs paying the higher “gift exchange” wage maximizes profits.

JEL Classification: C91, C93, J33, J41, D03

Keywords: gift exchange, reciprocity, multi-tasking, laboratory and field experiments

Corresponding author:

Robert Slonim
370 Merewether Building
Faculty of Economics and Business
University of Sydney
Sydney 2006
Australia
E-mail: roberts@econ.usyd.edu.au

* This research has benefited from discussions with Gary Becker, Dana Chandler, Ellen Garbarino, Pablo Guillen, Glenn Harrison, Edward Lazear, Steven Leider, John List and seminar participants at the University of Chicago, the University of Michigan, the University of Technology and Science (Sydney), the Fourth Annual Australian Workshop in Experimental Economics in Christchurch, NZ (2009) and the Asia Pacific Experimental Science Association Meetings in Melbourne, Australia (2010). The authors greatly appreciate the financial support from the Faculty of Economics and Business at the University of Sydney and the Australian Red Cross Blood Service. The physical surveys entered by workers in this experiment were developed and collected by Ashley Craig, Ellen Garbarino and Robert Slonim with support from the Australian Red Cross Blood Service. We also appreciate computational support from Sophia Chong, Ashley Craig and Malvin Sharma.

1. Introduction

Most jobs require workers to provide effort towards multiple outcomes, of which only a subset can be easily observed by the employer. Theoretical work on this “multitasking problem” (Holmstrom and Milgrom 1991; Baker, Gibbons and Murphy 1994) predicts that contracting on the easily observed outcomes can distort workers effort towards giving too little effort to the unobservable outcomes relative to the profit maximizing effort allocation.¹ This potential shirking on the harder to observe outcomes suggests that non-contractual solutions are necessary to improve worker productivity. One potential solution is that firms offer gift exchange wages (Akerlof 1982) which could motivate workers with reciprocal preferences.

Unfortunately, little is known empirically about whether reciprocal preferences will increase or decrease the amount of effort across all dimensions of work, across only observable or unobservable dimensions, or across no dimensions when workers need to multi-task. Reciprocal preferences on one clearly defined dimension of effort have been observed in many laboratory studies (e.g., Fehr, Kirchsteiger and Riedl 1993; see Charness, Frechette and Kagel 2004 for a review) and are increasingly used to model preferences across a broad range of contexts. Existing laboratory studies do not offer much evidence on generalizing this topic to multiple dimensions of work, however, since they have almost exclusively examined a single and clearly observable dimension of effort,² and indeed most laboratory studies have been explicitly designed to rule out any means to reciprocate other than through a single observable dimension.³

Recent field experiments (Gneezy and List 2006; Cohn, Fehr and Goette 2009; Kube Marechal and Puppe 2006; Hennig-Schmidt, Rockenbach and Sadrieh 2010; Kube et al 2011-wage treatment) have also examined whether workers reciprocate higher wages with greater effort. Similar to the laboratory experiments that preceded them, these field experiments (with the exception of Hennig-Schmidt et al.) also focus on a single and easily observable dimension of effort, in particular the quantity of work (e.g., data entered).⁴ None of these experiments finds that it is profitable to pay a higher wage,⁵ casting doubt on

¹ For a general review of the literature on the role of incentives in organizations, see Prendergast (1999)

² A notable exception is Fehr and Schmidt (2004). However, they do not study the gift exchange hypothesis; they study whether a principal’s ability to offer a bonus *after the work has been completed* induces greater effort on multiple dimensions.

³ The exclusive focus on a single dimension has been a positive feature of the lab studies for the intended purpose of the studies, only allowing subjects to express their reciprocal behavior through a single observable effort choice. However, this feature is at odds with understanding the current question of multi-dimensional effort allocation. Xiao and Houser (2005) offer an exception to the single dimension for reciprocation. In their ultimatum game experiment, subjects could send messages in addition to either accepting or rejecting offers. They find that fewer offers were rejected compared to when no message was permitted, indicating subjects were substituting between two methods to reciprocate offers.

⁴ Gneezy and List (2006) note that while the number of houses visited during a door-to-door fundraising task did not change when workers were paid a higher wage, the amount donated increased, suggesting workers may have reciprocated on a quality rather than quantity of visits. However, since their sample size was small and the workers’ behavior was not directly observed, a variety of other factors could have varied with the worker effort so that it is difficult to directly measure quality in their study.

⁵ Gneezy and List (2006) find that paying 67% higher wages led to only 10% more books being catalogued and paying 100% higher wages led to only 38% more charitable donations (it is possible that charitable donations may be higher in the long run, the reported 38% increase in donations was only on the initial approach). Kube et al. (2008) find that paying a 33% higher wage led to 10% (28%) more books catalogued when the higher (lower) wage was a surprise. Cohn et al. (2009) find over all their workers that paying a 33% higher wage resulted in only 3% more newspapers being handed out. Kube et al (2011) find that a 20% pay increase increases worker’s productivity by 5%; they also find that an in-kind gift of equivalent monetary value increased productivity by 25%.

the gift exchange hypothesis. Moreover, these studies find mixed results regarding whether workers even produce greater output when paid higher wages. However, these findings differ with other empirical literature that finds support for reciprocal behavior in labor market contexts (Mas 2006, Krueger and Mas 2004). Mas (2006) and Kreuger and Mas (2004), who focus on negative reciprocity, find that workers appear to reciprocate on quality dimensions of the work. The relative difficulty for the employer to observe one or more dimensions of work may thus create incentives for workers to shirk along the unobserved dimensions (Holmstrom and Milgrom 1991; Baker, et al. 1994), and may therefore also be the relevant margin for workers when they are reciprocating. Since many dimensions of work are unobservable, and thus not easily contractible, relying on a reciprocal response may be the firms' only means of increasing effort without (further) distorting the workers effort allocation across dimensions.

To examine the reciprocal responses when workers have multiple dimensions of effort, we ran a hybrid field-laboratory experiment. The experiment employed a 2 x 2 between-subjects design in which we manipulated the wage (High=\$22, Low=\$17) and the framing of the wage (Fair, Unfair). Our task is similar to the data entry field experiments (Gneezy and List 2006 & Kube et al. 2008) in that we advertised for a two-day data-entry job and had workers enter genuine data for an organization that had intrinsic value for the data and paid for the work, but the work was performed in the laboratory as an experiment and workers were aware they were entering data while simultaneously participating in an experiment.

While we do not find any significant reciprocation in terms of the total number of surveys entered, we detect significant reciprocation with workers in the higher wage treatment producing consistently higher quality work, which translates to higher quantities of output produced if quality restrictions are imposed on the output. This provides one reason why firms may profitably offer gift-exchange wages; to ensure quality, the cost of offering a higher wage may be less than the costs of alternative methods such as monitoring. We find that if the employer places a value v on each error that occurs in the data that is \$0.89 or higher, then it would have been profitable to offer the higher wage; the number of surveys entered per dollar (including the costs of errors) was higher in the high wage condition. Thus, as the value of quality increases, the higher wage becomes relatively, and for $v > \$0.89$, absolutely, more profitable.

The current research also contributes several innovations to the methodology of studying reciprocity and the gift exchange hypothesis in labor markets. First, like Al-Ubaydli et al (2008) we study both the quantity and quality of effort. When we examine the quantity produced, our results are similar to those reported in the field studies looking at just one dimension of effort; we find that paying a high wage only has a small effect on the total number of surveys entered. However, when we examine the quality of the work, we find significant reciprocation.

The second innovation is that the laboratory element of the experiment provides more control than the previous field work. Other than the two manipulations, the wage and the framing of the wage, the data that was entered and the environment in which it was entered were identical in every respect between the treatments. In particular, we were able to control for the difficulty of the data-entry work and also measure

each worker's productivity when paid a high piece wage in order to obtain a proxy for each workers maximum level of effort. Given the relatively small sample sizes and potentially high heterogeneity in worker ability, our measure of maximum effort, as designed, substantially improves the precision of our treatment effect estimates. Combined, the greater control and measure of maximum effort substantially increases (by nearly 50 percent) the precision with which we estimate treatment effects; without this added precision, some estimates would not have reached the level of significance.

The design also allows us to collect perceptions of wages before and after the work was done to estimate the treatment effects on workers perceptions of the fairness of the wage and whether they persist or change over time. We find that both the wage and framing manipulations significantly affected wage fairness perceptions and these perceptions persisted over both days. However, we could not detect any significant effect of the framing manipulation on either quantity or quality of output, suggesting that the wage manipulation may be working through some other channel than perceived fairness.

Finally, the laboratory let us stress the importance of the work to the employer. Reciprocal responses may be strongly dependent on the nature of the work, both on the marginal benefit to the employer who offers the higher wage, and on the ability of increased effort to boost productivity (Englmaier and Leider 2009, 2011; Hennig-Schmidt, Rockenbach and Sadrieh 2010); the closer the relationship between effort and productivity and productivity and the benefit to the employer who offers the gift, the greater the channel for reciprocity. In the lab, we will explicitly stress these relationships.

The paper is organized as follows. Section 2 describes the hybrid field-laboratory experiment. Section 3 presents the results, starting with estimates of treatment effects on perceptions of wage fairness and then analyzes the treatment effects on (1) quantity, (2) quality and (3) quantity adjusting for quality. We also present results on the importance of the lab controls and conclude by exploring whether it was profitable to pay the higher wage after taking into account marginal costs associated with errors. Section 4 summarizes the results and their implications for our understanding of reciprocity and the viability of the gift exchange hypothesis in labor markets.

2. Experimental Design

2.1 Instructions and Context

The experiment was run at the University of Sydney. The task was a one-time, two-day data-entry job for a well-known large non-profit organization.⁶ We stressed to the workers that the data would be used to help the organization. We also stressed that there was no opportunity for further employment beyond the two days of work in order to control for reputation effects.

We recruited workers from a list of students who had expressed interest in participating in experiments. Our recruitment e-mail indicated that participants would earn between 15 and 29 (Australian) dollars per

⁶ Gneezy and List (2006) find significant reciprocal responses in both a book-cataloguing and donation solicitation task, but the reciprocity was only sustained for few hours. Thus, as Al-Ubaydli et al. (2008) argue, a two-day job serves as a robustness test.

hour⁷ for the data entry work, and that they had to attend two 2.5-hour days of work one week apart in order to be paid. We recruited 21 workers for each two-day session, and hired the first sixteen workers who arrived on Day 1, dismissing the remainder with a \$10 show-up fee.⁸ The experiment consisted of a 2 x 2 between-subjects design which manipulated the wage (\$17 and \$22 an hour) and the framing of the wage paid (Fair and Unfair). The two days of data entry for each treatment were one week apart and ran from 1:15 to 3:45 PM.

Each worker was randomly assigned to one of sixteen computer workstations based on the order they arrived.⁹ The workstations had partitions such that workers could see the monitors in the front of the room but could not see other workers, and were seated two desks apart from each other to minimize workers being able to hear each other working. This set up was designed to minimize peer effects.¹⁰ Nonetheless, we wanted many workers in the room at one time to reflect common data entry work environments (Al-Ubaydli et al (2008) and Gneezy and List (2006) use a similar set-up in their field experiments) and used the partitions to minimize worker interaction in order to maintain independent experimental observation. Each treatment was filled to capacity (16 workers) each day except for one worker who arrived too late on the second day to participate.¹¹ Two monitors were always present to assist the workers with data entry questions and to ensure that workers did not communicate with each other.

All sessions were identical except for the wage and framing manipulations. At the beginning of the first day, each worker was given a written set of instructions. A monitor read them aloud and asked the workers to follow along. The instructions began by emphasizing the importance of the work by describing the charitable organization and a major problem that the organization was trying to solve. The instructions then described the data entry process and the specific procedures to follow when common problems were encountered (e.g., response illegibility, spelling mistakes and multiple responses when only one response was requested). After these instructions workers were asked to enter two “made-up” sample surveys to give each worker some experience with the work that they would be doing. One of the sample surveys included many of the common problems they could expect to encounter in order to familiarise the workers with the issues that they would be dealing with. Monitors were available to answer questions during this time to ensure that each worker had a comprehensive understanding of the data-entry process.

2.2 The Data Entry

⁷ These are Australian Dollars. At the time of the experiment, \$1.00 U.S. was worth approximately \$1.20 Australian, thus the advertised wage in U.S. dollars ranged from \$12.50 to \$24.17 in US dollars.

⁸ We sent the e-mail to a randomly selected group of 200 students. The e-mail contained a link to a website that provided the dates and times of the four two-day sessions. All sessions were filled to capacity in less than 24 hours. A reminder e-mail was sent the night before each session to everyone who was signed up for that session.

⁹ We predetermined where each participant would sit based on the order they arrived and we used the same order for each treatment. In addition to the participants being separated over one meter apart, participants were never seated immediately next to, in front of or behind the person who arrived at the experiment just before or after them.

¹⁰ This was done in order to minimise peer effects, which may compromise the independence of the observations: workers may work harder (or less hard) if they can see how hard others are working or know that others can observe how hard they are working. Mas and Moretti (2009) find that grocery store workers' speed at checkout counters is significantly affected by the speed of other workers they can see.

¹¹ The subject came over 60 minutes into the data entry period on Day 2 and explained that she was late because she had to make up a class. She was dismissed from the study when she arrived and we therefore have no data on this worker for Day 2.

The job required workers to enter data from physical (paper) surveys into a web-based electronic form. The web-form was designed to look identical to the physical surveys, with the exception of two text-entry boxes not on the paper version, one to enter notes left by the survey responders and one to enter notes left by the research staff who administered the surveys. Every survey included 51 radio-button questions (each one asked for one response), 20 questions requiring numerical responses and five questions requiring open ended written responses.

The surveys were randomly divided into 16 “packets,” or groups of distinct surveys, which were then linked to a specific physical workstation. In each treatment, each of the 16 workers was uniquely assigned to one of the 16 packets. Since all packets contained unique surveys, we truthfully informed the workers that they were entering different surveys than any other worker who was in the room with them.¹² However, across treatments the packets were identical in every respect; thus the worker who was given Packet 1 in each treatment would be given the identical set of surveys and in the identical order and would be seated at the identical workstation location. By providing the identical surveys in the identical order across treatments, we control for the difficulty of data entry (e.g., harder or easier to read text, variation in the length of text to enter, hand-writing legibility, etc.). By placing each packet in the identical workstation location across treatments (e.g., Packet 1 was on the far left end of the 3rd row), we further diminish variation across output for each packet that could be due to location effects (e.g., closer or further from the aisle, in the front or last row, etc.). Our packet design provides especially strong control for the potential variation in the difficulty of the work across treatments. Each packet contained 60 surveys on the first day and 70 surveys on the second day, far more than any worker completed.¹³

2.3 The Manipulations

After the practice surveys were entered, we asked each worker to read and sign an employment contract. The contract was read aloud and differed between the four treatments: (a) Unfair Framing, \$17 Wage, (b) Fair Framing, \$17 Wage, (c) Unfair Framing, \$22 Wage, (d) Fair Framing, \$22 Wage. This one-page contract constituted the bulk of the manipulation, with one additional framing manipulation during the personal survey (see below).

The contract was divided into three parts. Part One established the wage and was identical in all treatments other than the wage:

“You will be paid \$17 (\$22) per hour for the entire duration of the five hours that you spend here, including the time you spend reading the instructions, filling out surveys and learning to do the work. This means that you will receive:

\$17(\$22) per hour for five hours equals \$85 (\$110) in total.

Please keep in mind though that no payments will be made if you do not return and complete next week's data entry session unless there is an emergency that prevents you from attending next week.

¹² They were not told that the surveys would also be entered on other days.

¹³ The average number of surveys completed per worker on Day 1 and Day 2 was 27.5 and 30.3, respectively, and the maximum was 39 and 46.

Thank you for helping us do this important work as efficiently as possible.”

Part Two framed the wage as either fair or unfair and thus the language varied substantially between treatments.

Fair framing:

“Please note that this wage is relatively high compared to the New South Wales state minimum wage that ranges from \$9.78 to \$17.20 an hour for clerical work of this type. Moreover, please keep in mind that the [organization name] is an organization that relies on donations for its existence, making budgets tight for research projects such as this one. Finally, note that although we advertised paying as little as \$15 per hour, we are paying some participants a little more due to the importance of this work. The total sum paid of \$85(\$110) for this experiment is well above the average typically paid to participants in other experiments. This research is important to the long-term goals of the [organization name] and the effectiveness of their operation in meeting [specific to organization] donation goals, and ultimately helping save lives.”

Unfair framing:

“Please note that the university mandate exempts laboratory experiments from the minimum wage ranging from \$23 to \$36 for clerical work such as data-entry. Although we advertised paying as much as \$29 per hour and some other experiments typically pay \$25 per hour or more on average to participants, we are paying some participants a little less because of limited budgets for this work as this will allow us to get more data entered for the same budget.”

The framing aimed to induce a perception of the wage as Fair (Unfair) on four dimensions:

1. *Wages for similar work:* We favorably compared their wage with the lower state minimum wage in the fair framing and unfavorably compared their wage with the higher university minimum wage in the unfair framing.
2. *Expectations:* We favorably compared their wage with the lower bound of the range advertised in the fair framing and unfavourably compared their wage with the upper bound advertised in the unfair framing.¹⁴
3. *Wages for other experiments:* We favorably compared their total earnings in the fair framing to *total* earnings in other experiments and unfavorably compared their hourly earnings to other experiments *hourly* earnings.
4. *The employer’s budget:* We justified their wage in the fair treatment by stating that the researchers were working with a tight budget. In contrast, in the unfair framing we strongly imply that the purpose of the experiment was to circumvent the university minimum wage.

We also reiterated the intrinsic value of the work in the fair framing to increase the perception of the work’s importance to the charity organisation.

Part Three of the contract asked workers to sign their contracts. We required workers to sign their contracts to reinforce that this was meaningful work.

Once workers signed their contracts, they were asked to complete a survey that took between 10 and 15 minutes. This survey collected information on factors that could affect productivity, including

¹⁴ Al-Ubaydli et al. (2008) use a similar expectations manipulation in their surprise treatments; they advertised a wage range between \$8 and \$16 per hour, and then paid workers either \$8 (negative surprise) or \$16 (positive surprise).

demographics, perceptions of the work, familiarity with, and perception of, the charity, previous work experience and their current mental and physical condition.

The survey also included one additional manipulation of the perception of the wage. We asked all workers the following question in order to assess their perceptions of fairness for not only their own wage, but also other possible wages:

“Although we are paying you \$17(\$22) per hour, we would like to know how fair you would think other possible wages might have been. For each wage listed, please click on the response, from Extremely Unfair (1) to Extremely Fair (7) that best reflects how you would feel about getting paid each wage for the data entry task you will be doing today and next week. “

We manipulated the order and range of wages in which to answer this question as follows:

| Survey Question | Day 1 Survey (before work) Treatment Conditions | | | | Day 2 Survey (after work) |
|-----------------|--|--------------------|------------------|--------------------|---------------------------|
| | <u>Order</u> | <u>Unfair \$17</u> | <u>Fair \$17</u> | <u>Unfair \$22</u> | <u>Fair \$22</u> |
| 1 | \$26 | \$8 | \$31 | \$13 | \$15 |
| 2 | \$25 | \$9 | \$30 | \$14 | \$16 |
| 3 | \$24 | \$10 | \$29 | \$15 | <u>\$17</u> |
| 4 | \$23 | \$11 | \$28 | \$16 | \$18 |
| 5 | \$22 | \$12 | \$27 | \$17 | \$19 |
| 6 | \$21 | \$13 | \$26 | \$18 | \$20 |
| 7 | \$20 | \$14 | \$25 | \$19 | \$21 |
| 8 | \$19 | \$15 | \$24 | \$20 | <u>\$22</u> |
| 9 | \$18 | \$16 | \$23 | \$21 | \$23 |
| 10 | <u>\$17</u> | <u>\$17</u> | <u>\$22</u> | <u>\$22</u> | \$24 |
| 11 | \$16 | \$18 | \$21 | \$23 | \$25 |

The orders and ranges suggest a mental context for the relative fairness of the wage the workers were getting paid. We placed their actual wage near the high (low) end of the range in the Fair (Unfair) treatments. In all treatments, the wage each worker was paid appeared in the next to the last question (Question 10 above) after they had seen and answered the proceeding nine questions with lower (higher) wages in the Fair (Unfair) condition. The expectation is that the rising wage sequence in the Fair framings will lead to the same actual wage (\$17 or \$22) being perceived as fairer than the falling wage sequence in the Unfair framing.¹⁵

The instructions, practice data entry and survey took almost 50 minutes. Workers then entered data for 100 minutes. During this time they worked almost entirely autonomously, occasionally raising their hands to ask the monitors a question. The monitors announced when 50 and 98 minutes had passed. When 100 minutes passed, the monitors asked the workers to finish the survey they were currently working on. This completed the first day of work.

¹⁵ Slonim and Garbarino (1999) examine a similar price sequence manipulation. They find that subjects perceived the price of a good to be relatively more expensive (and less fair) after observing lower prices than observing higher prices, ceteris paribus.

All workers (except one) returned exactly one week later and were seated at the same workstations they sat at the first day. The instructions and all procedures on Day 2 were identical across all treatments. There was no reminder of the framing at any time during Day 2, and no reminder of the wage until workers were paid after all work had been completed. Day 2 began with a monitor reading a short set of instructions to refresh the workers on the data-entry process, which took just under 5 minutes. The workers then entered data for 100 minutes. All procedures and conditions during the 100 minute data entry time on Day 2 were identical to Day 1. After the 100 minutes, the workers answered another personal survey which took on average five minutes to complete. This survey contained a subset of questions from the first survey, including questions on their physical condition and their perception of the wage. The range of wages on this survey was identical across all four treatments and went from \$15 to \$25 (see insert above). This survey also asked a few control questions regarding their beliefs in the authenticity of the work and whether they communicated with anyone else in the experiment between Day 1 and 2.¹⁶

2.4 Inducing workers to give maximum effort

After completing the Day 2 survey, workers were instructed that there was one more task that would take 30 minutes. For this *piece-rate* task, we prepared 30 surveys for each worker's packet. These surveys were screened to be approximately equally difficult to enter across the 16 packets by filtering out those surveys which had an unusually large amount of text or difficult entries. We informed the workers that in addition to their fixed wage, during the final 30 minutes they would earn an additional \$2 per survey completed multiplied by the percent of correct entries they made; thus they had a strong incentive to enter surveys quickly and accurately. Workers on average earned \$23.10 for the 30 minutes of piece rate work, equivalent to an hourly rate of \$46.20, a substantially higher wage than available to the vast majority of students.

We chose a very high piece rate wage in order to provide a substantial incentive for the workers to work as hard as they possibly could. Since we multiply each survey in the half-hour session by two to determine their earnings, the payment in the piece rate session can be interpreted as the per-hour rate of entry during the session, adjusted for the error percent, and can be interpreted directly as an approximate measurement of the workers productivity when they give maximum effort. To the extent that all else is equal between the fixed rate and piece rate data entry sessions (e.g., worker's ability, preferences toward the organization and room location), the difference in productivity between a worker in the fixed rate and in the piece rate sessions can be attributed to differences in effort caused by the treatments. Thus, controlling for productivity during the piece rate session should explain variation in output during the fixed rate times due to variation across workers due to ability, preferences, and other unobservable individual differences that could affect productivity. Last, we chose to include the piece rate task after the completion of the main experiment with the fixed wage in order to minimize any effect it might have had on the workers. For instance, it could potentially have changed all subjects' perception of their fixed rate

¹⁶ Only two subjects indicated they had talked with anyone about the work between sessions.

condition as well as affected the speed at which the workers learned to enter data that might otherwise vary with the main manipulations.

2.5 The Workers

Table 1 describes the workers; 43 percent (27/63) were female, 43 percent had experience entering data, 49 percent were currently employed, 41 percent were in the Economics and Business faculty and, on average, they were 21.8 years old. None of these characteristics differ significantly between the four treatments. Since the work was advertised through an experimental subject population, we stressed to the workers in the instructions that they were doing authentic work while participating in an experiment. To examine whether they believed they were doing authentic work, in the final survey we asked, “Did you doubt any aspect of the experiment?” Over 90 percent (57/63) indicated no doubt. Nonetheless, we control for doubt throughout the analysis since those who expressed doubt may have given less effort.¹⁷

Table 1 also shows the mean payment to each worker in the piece rate condition. Under the identical piece-rate conditions, workers in Unfair \$22 earned the most in the piece-rate session while workers in Unfair \$17 and Fair \$22 earned the least; workers in Fair \$22 earned five and ten percent less than workers in Fair \$17 and Unfair \$22, respectively. Figure 1 shows the distribution of earnings in the piece rate session for each treatment. Although the means across the treatments differ by 10 percent or less, Figure 1 shows substantial heterogeneity across individual workers; for instance, nearly 20% (12/63) of the workers earned over \$27 in the piece rate session while 16% (10/63) earned under \$18. We show below that controlling for this heterogeneity improves the precision of the estimates of the treatment effects substantially.

3. Results

We first show (3.1) that the Fairness and Wage treatments altered worker’s perception of wages as intended throughout both days of work. Section 3.2 defines our measure of the *quantity* of output. We then examine the effect of the treatments on the total number of surveys completed ignoring quality (3.3) and on the quality (accuracy of entries) of surveys completed ignoring quantity (3.4). We initially explore each dimension of effort independently since both quantity and quality were stressed in the instructions, and since we anticipate that the response to the treatments may differ due to the greater perceived observability of quantity than quality. Section 3.5 then examines the treatment effects on a quantity-quality aggregate measure of effort by adjusting the number of surveys completed for quality. Section 3.6 considers whether the reciprocity induced by the higher wage was sufficient to make paying the higher wage profitable, a result upon which the viability of the gift-exchange hypothesis depends. Last, section 3.7 shows the advantages of the controls gained from conducting this work in the lab.

3.1 Treatment effects on wage perceptions

¹⁷ We do not detect statistical differences ($p > .15$) between those who did and did not indicate doubt. The mean number of surveys completed among the six workers who expressed doubt during Day 1 and Day 2 was 15.4 and 18.2, respectively. The mean number among workers expressing no doubt was 16.8 and 18.5.

We find that both the Framing and Wage treatments significantly influenced workers perception of the wage. Table 2 presents regression estimates of the effect of the manipulations on fairness perceptions where the dependent variable is the workers perception of fairness as reported in the personal surveys.¹⁸ The surveys were completed by all workers at the beginning of Day 1 before any work had started, and on Day 2 after the fixed wage time of the work had been completed. The potential responses range from 1 (Extremely Unfair) to 7 (Extremely Fair). Figures 2a and 2b present the distribution of wage perceptions by treatment on Day 1 and Day 2, respectively. These figures show that there is a general shift across the distribution to perceive the wage as fairer in the Fair than Unfair conditions and in the High than Low Wage conditions.

Columns 1-3 in Table 2 show estimates from Tobit Regressions (censored at 1 and 7), and Columns 4-6 present estimates from ordered Probit Regressions.¹⁹ The estimates in all of the regressions show that the High Wage and the Fair Framing significantly ($p < .01$) increased the perceived fairness of the wage on Day 1 (Columns 1 and 4) and Day 2 (Columns 2 and 5). Columns 3 and 6 stack the data from both days and add interaction terms for Day 2 by Fair Framing and Day 2 by High Wage. Since there are two observations per worker, one for each day, we estimate robust standard errors clustering on workers. Neither interaction is significant ($p > .20$), indicating that the effect of the two manipulations on the perception of the fairness of the wage did not diminish significantly. We also examined whether there was an interaction effect between the High Wage and Fair Framing treatments on wage perceptions, but found no significant effect ($p > .20$) on Day 1, Day 2 or across both days combined.

The estimates in Table 2 and Figures 2a and 2b show that Unfair \$17 workers perceived their wage to be on average below the median fairness level whereas Fair \$22 workers perceived their wage to be, on average, almost the highest possible fairness value. The estimated wage perception among Fair \$17 and Unfair \$22 workers was almost exactly half way between Unfair \$17 and Fair \$22 worker perceptions, suggesting that the two manipulations had a similar effect on fairness perceptions. Fairness perceptions were maximised when the higher wage was framed as being Fair – it was not sufficient to maximize perceptions by simply paying the higher wage.

Result 1: The High wage and Fair framing treatments significantly increased workers fairness perception of the wage. The effect persisted over two working days one week apart.

3.2 Measuring Productivity

The first measure of productivity y_{it} we examine is the total number of surveys completed by worker i during time interval t converted to a per hour rate. We divide the data into nine ten minute work intervals

¹⁸ The question is presented in section 2.3 above. We use the response to the wage they were paid as our dependent variable. Although the ranges for this question formed part of the manipulation on the Day 1 survey, the range on the Day 2 survey was identical for all treatments.

¹⁹ One worker did not provide a response to any of the wage perception questions on Day 1 or Day 2. However, this worker provided a response to every other question on both surveys. This subject asked a monitor how to define “fair” – the monitor instructed the subject to use his best judgment. The question of how the workers interpret “fairness” and whether this corresponds to the notion of “fairness” used by economists has not been explored in depth to our knowledge.

(t) each day to control for non-linear changes in output over time within each day.²⁰ The first ten minute interval for each worker each day begins at the time the worker completed their first survey. We do not include the first survey in our analyses since there is variability in when it was completed which is unrelated to worker effort. Nonetheless, all workers completed their first survey within the first eight minutes (and almost all within the first six minutes) of the 100 minute period. This means that the last (ninth) ten-minute interval we use in our analysis ends two to seven minutes before the end of the 100 minute period.

The data entry software recorded the time (to the second) that each survey was submitted, allowing us to construct a nearly continuous measure of productivity. For each 10 minute interval we determine productivity by summing all the surveys completed entirely in the interval plus the fraction that overlap the beginning and end of the interval.²¹ Following this procedure, we construct a measure of the quantity of surveys completed for each worker for 18 ($t = 1$ to 9 for Day 1 and $t = 10$ to 18 for Day 2) observations.

Figure 3a graphs the average number of surveys per hour completed by each worker for each treatment during the nine time intervals on Day 1 and Day 2. Figure 3 shows that within each day workers are completing surveys faster over time, workers completed more surveys during Day 2 than Day 1 and workers began Day 2 at a rate slightly slower than where they ended on Day 1.

3.3 Quantity Produced (ignoring quality)

Figure 4 presents the mean number of surveys completed in each treatment for each day. On both days surveys completed is higher in the Fair \$17 than the Unfair \$17 treatment and likewise is higher in the Unfair \$22 than Unfair \$17 treatment, directionally consistent with reciprocity. In contrast, Figure 4 also shows that the mean number of surveys completed is lower in Fair \$22 than Unfair \$22 and likewise is lower in Fair \$22 than Fair \$17. However, the average measures may be misleading for at least two reasons: (1) no control for individual differences, especially in ability (recall from Table 1 that workers in the Fair \$22 had the lowest piece rate productivity), and (2) no adjustment for the quality of the data-entry work. To test whether total surveys completed differs *significantly* across treatments, we estimate the following OLS regression model:

$$(1) \quad y_{it} = \alpha + \beta_1 * \text{Fair}_i + \beta_2 * \text{High Wage}_i + \beta_3 * \text{Ave Chars}_{it} + \beta_4 * \text{Piece Rate Output}_i + \sum \theta_j * X_{ij} + \sum \rho_k * P_{ki} + \sum \delta_t D_t + \varepsilon_i$$

where y_{it} is the hourly rate of surveys completed by worker i during time t , Fair and High Wage are dummy variables that equal 1 if worker i is in the Fair Framing or High Wage treatments, respectively, Ave Chars is the average number of text characters per survey the worker completed during time t ; Ave Chars provides an approximate measure of survey-specific difficulty a worker encountered during each

²⁰ Al-Ubaydli et al. (2008), Englmaier and Leider (2009) and Gneezy and List (2006) similarly divide their data into distinct time intervals.

²¹ For instance, suppose worker X completed surveys at 37, 41, 45, 48 and 52 minutes. In this case, the number of surveys completed for the 40 to 50 minute interval equals the two surveys completed entirely (at 45 and 48) plus one-quarter of the survey completed at the beginning of the interval ($[41-40] / [41-37]$) plus one-half of the survey completed at the end of the interval ($[50-48] / [52-48]$) for a total of 2.75 surveys. We then convert this sum to surveys per hour (in this case, $16.5 = 6 \text{ times } 2.75$).

time interval, and is anticipated to be inversely related to the quantity of surveys completed. Piece Rate Output is the workers productivity during the 30 minute piece rate time, X_{ij} is a vector of covariates that includes the survey measures of doubt about the authenticity of the experiment, how well workers knew the charity organization, and how fatigued or tired workers were on the day they did the data entry work. P_{ki} is a vector of dummy variables for each packet k , D_t is a dummy variable for time interval t ($t=1$ for 0-10 minutes on Day 1, $t=2$ for 10-20 minutes on Day 1, ..., $t=10$ for 0-10 minutes on Day 2, ..., $t=18$ for 80-90 minutes on Day 2), and ε_i is the error term. Since we have 18 observations for each worker we estimate robust standard errors clustered by worker.²²

Table 3 presents estimates of the treatment effects on the total number of surveys completed. Column 1 shows the average effect of each treatment. Although the Fair Framing and High Wage treatments both increased productivity directionally, neither coefficient is significant ($p > .20$). Column 2 adds controls for time, packets and the survey measures. With these controls, the point estimates for both treatments remain positive but again neither coefficient reaches a conventional level of significance. However, these controls improve the precision of the point estimates on the treatment variables; the standard errors on both treatment estimates falls by approximately 14 percent and the model fit (R^2) improves substantially. Column 3 adds control for the Average Characters on each survey completed. As anticipated, surveys with more characters caused workers to complete significantly fewer surveys ($p < .001$); on average, for each additional character per survey, workers completed 0.033 fewer surveys per hour. Finally, note that controlling for Average Characters further increased the model fit substantially from $R^2 = .235$ to $.334$, though it had a small negative effect on the precision of estimating the treatment effects.

Column 4 adds control for each worker's piece-rate output. The piece-rate point estimate of 0.78 implies that on average workers under the fixed wage scheme were completing surveys at a rate that was 78 percent of their maximum effort. This point estimate is highly significant ($p < .001$).²³ Moreover, controlling for piece-rate output improves the precision of the treatment effect estimates by an additional 40 percent and the model fitness (R^2) again increases substantially. Despite this increased precision, the estimated treatment effects remain insignificant, although they remain positive and the t-statistics are now greater than 1.

Column 5 examines whether the treatment effects are different across the two days. Although neither point estimate for the interaction terms reaches the level of significance, the directionally negative estimate for the interaction of Day 2 by High Wage suggests that reciprocity could be decaying over time.

²² Other specifications, such as an interaction term between the treatments, were estimated but had no qualitative effect on the results. We also analyzed the total number of characters as our measure of quantity and obtain qualitatively similar, but noisier estimates. However, using the number of characters introduces a number of ambiguous factors, such as how to treat judgment on what was intended on the surveys, and was highly variable across surveys. We thus report the estimates for number of surveys as our quantitative measure of output since we felt it was the more direct and unambiguous measure of output, and perhaps more importantly was the measure of output the workers would most likely assume would be observable (see discussion below).

²³ The piece rate estimate is also significantly less than 1 ($p < .01$) indicating workers were less productive during the fixed than piece rate conditions. This could be due to several factors including exhaustion at the end of the fixed rate period, learning or shirking. However, exhaustion seems unlikely since the piece rate period came immediately after the fixed rate time. Learning also seems unlikely; Figure 3 shows productivity remained fairly constant during the last hour of the fixed rate period. We thus interpret the piece rate estimate to reflect significant shirking.

The estimates in Column 5 also show that the main effect of offering the High Wage increased the total number of surveys completed on Day 1 alone by 0.77 surveys per hour; although this estimate does not reach a conventional level of significance ($p=.11$). Considering that 17.5 surveys were completed per hour on average, paying the higher wage during Day 1 increased productivity directionally by approximately five percent.²⁴

To this point in the analysis, our results are largely consistent with the existing experimental field studies examining the effects of wages on total quantity produced. We find that the higher wage led to directionally greater productivity during Day 1 but not during Day 2, similar to Gneezy and List's (2006) evidence that a higher wage for books catalogued increased productivity during the first 1.5 hours of work, but had no significant effect thereafter. Likewise, Cohn et al (2009) found a positive but insignificant effect of the higher wage on productivity in similar regressions that did not control for other factors.

Result 2: Ignoring quality, neither the High Wage nor the Fair Framing significantly affected the total number of surveys completed over both days combined or either day alone.

3.4 Quality Produced (ignoring quantity)

We now investigate whether workers were providing different levels of effort towards the quality of their work. Specifically, more effort could result in workers making fewer errors coding the surveys. We restrict the analysis of errors to the 51 radio button responses on each survey.²⁵ Figure 5 shows the average percentage of errors for each day and treatment. It shows that on each day workers on average made fewer errors in the High than Low Wage condition. On Day 1 workers also made fewer errors in the Fair than Unfair conditions, whereas on Day 2 workers made slightly fewer errors in the Unfair than Fair conditions. To examine whether the treatments affected quality significantly, we estimate the following Tobit model censoring the dependent variable (percent errors) on the minimum possible error rate of 0 percent):

$$(2) \text{ Perc Errors}_{it} = \alpha + \beta_1 * \text{Fair}_i + \beta_2 * \text{HighWage}_i + \beta_3 * \text{Ave Chars}_{it} + \beta_4 * \text{Piece Rate Output}_i + \sum \theta_j * X_{ij} + \sum \rho_k P_{ki} + \sum \delta_t D_t + \varepsilon_i$$

where Perc Errors_{it} is the percent of errors worker i made during time interval t . We again estimate standard errors clustered by worker.

Table 4 presents the results. Column 1 shows that workers who were paid the higher wage on average made 0.59 percent absolute fewer errors than workers paid the low wage. This difference is significant at the 5 percent level. Given that workers in the two low wage conditions on average made errors 2.28% of the time, a 0.59% absolute reduction in errors translates to a 25.9% ($=0.59/2.28$) reduction when compared with the Low Wage condition. Column 1 also shows that on average workers in the Fair condition did not make significantly more or less errors than workers in the Unfair condition. When we include the control

²⁴ Regressions analysing the data from Day 1 and Day 2 separately using the model in Column 4 show that the High Wage marginally significantly increased output on Day 1 ($p<.10$) and had no significant effect on Day 2 ($p>.30$).

²⁵ We only consider radio buttons for the quality analysis because there is no ambiguity as to the correct response. The numerical and free text entries include many subjective judgments regarding the correct response and are thus omitted from this analysis.

variables (Column 2), the estimated reduction in errors in the High Wage condition remains significant at the 5 percent level, the absolute decrease in errors is 0.55%, a relative improvement of 24.1% ($=0.55/2.28$). Column 3 shows that the average number of characters per survey has no effect on the percentage of errors, and no effect on the estimated treatment effects. Adding piece rate output to the model (Column 4), the estimated reduction in errors in the High Wage condition again remains significant at the 5 percent level, the absolute decrease in errors is 0.56%, and the relative improvement is 24.6%.

Column 4 also shows that the estimated effect of piece rate output on errors is 0.07 and significant ($p < .05$), indicating that workers who worked faster in the piece-rate condition were making more errors during the fixed rate condition. This is not surprising given we showed previously that workers completing more surveys in the piece rate condition were working faster in the fixed rate condition, and thus, *ceteris paribus*, would be expected to make more errors in the fixed rate condition.

The estimates in Column 5 indicate that on Day 1 workers in the high wage condition made 0.44% less errors than workers in the low wage condition and 0.67% fewer errors during Day 2. However the difference in errors made by high wage workers on Day 1 than Day 2 relative to low wage workers is not significant ($p > .20$). Last, Column 5 shows that workers in the Fair Framing made significantly more errors on Day 2 than Day 1. Regressions analysing data for each day separately (not shown) show that the Fair Framing did not significantly affect quality on either day alone, indicating that while the Framing effect may have differed across days, the effect of the Framing was not critical on either day alone.

Result 3: Workers made significantly fewer errors when paid the high than low wage and the difference persisted over both days. The Fair Framing did not significantly affect quality overall, but led to significantly more errors on Day 2 than Day 1.

Discussion: We anticipated that workers would be more likely reciprocate the higher wage on the dimension of the less unobservable quality dimension of work than on the more observable quantity of work dimension. In our work environment, however, we never mentioned that either the quantity or quality of the work would be monitored or observed (similar to the Al-Ubaydli et al (2008) field experiment). Nonetheless, since workers were entering surveys one at a time and submitting each one upon completion, it would have been readily apparent that the number of surveys could easily be observed by the software (or by our counting how many surveys were placed in each worker's completed box). In contrast, it was much less apparent that we could or would monitor quality, and thus quality should have appeared to workers as less observable than quantity. However, to the extent that any workers anticipated that we would monitor quality, and thus did not favour the quality dimension for reciprocating the higher wage, this could have caused us to under-estimate the effect of the wage treatment on increasing output towards the less observable quality measure.

3.5 Quantity Produced *Adjusted* for Quality

We now examine a measure of effort that accounts for both quantity and quality. Specifically, we examine whether the treatments affected the quantity of surveys completed given a quality threshold for a survey

being acceptable. One reason this is important to examine is that it provides a more complete single measure of total effort compared to quantity or quality alone. A second reason is that it reflects work places in which firms have implicit or explicit standards of quality, with many tasks involving a qualitative component that is likely to be more difficult to monitor than the number of units produced. For instance, suppliers are typically required to meet quality thresholds otherwise buyers can impose penalties or possibly cancel contracts, and the quality of the final product depends on the effort of many workers where it might not always be efficient or possible to monitor the quality of individual workers.

To examine quality adjusted output, we define a survey as completed ($y_{it}=1$) if and only if it is completed and meets a minimum quality threshold. We consider three thresholds. The first requires a survey to have less than five percent of the entries with errors. This restriction results in recording a survey as being produced if it has less than three errors (out of the 51 radio buttons); 89 percent of all surveys completed pass this restriction. The second and more restrictive threshold requires surveys to have less than two percent of the entries with errors. This restriction results in recording a survey as being produced if it has less than two errors; 77 percent of all surveys completed pass this restriction. The final and most restrictive threshold requires surveys to have no errors; 40 percent of all surveys completed pass this restriction.

Figures 3b and 6 highlight how quality restrictions affect productivity differentially across the treatments. Figures 6a and 6b present the average number of surveys that pass the three thresholds, completed on Day 1 and Day 2, respectively, and include for comparison all surveys completed (i.e., no quality restrictions). Figure 6 shows that on both days the wage effect is unambiguous once adjustments are made to control for quality. Recall that when we examined all surveys, directionally more surveys were completed each day in the Fair \$17 than the Fair \$22 treatment, inconsistent with workers reciprocating the higher wage. However, for each of the three quality thresholds on both days, Figure 6 shows that more surveys were completed in both Fair and Unfair \$22 than in Fair or Unfair \$17. Moreover Figure 3b, displaying the mean number of surveys per condition per ten minute intervals conditional on the 2% quality restriction, shows that across both days in all but two ten minute intervals more surveys were completed in the Fair \$22 than Fair \$17 and in all but one ten minute interval more surveys were completed in the Unfair \$22 than Unfair \$17. Thus, adjusting for quality, at least directionally, more surveys were completed in High Wage than Low Wage conditions consistent with a positive reciprocal response to the High Wage. However, the Fair Framing manipulation continues to have mixed effects on output after adjusting for quality.

Table 5 presents OLS regressions estimating Model 1 for the three quality adjusted definitions of surveys completed. For comparison, the first two columns show the results with no quality restrictions. The remaining columns show estimates for the increasingly restrictive quality requirements. For each restriction, we show the estimates with and without interactions for Day 2 by each treatment.

Columns 3-8 show that workers in the high wage condition entered significantly more surveys when we control for quality. Columns 3, 5 and 7 show that when there is a 5%, 2% or 0% restriction on

acceptable quality, respectively, 1.1 ($p < .05$), 1.6 ($p < .05$) and 1.5 ($p < .01$) more surveys per hour were completed in the High than Low Wage conditions. These differences are equivalent to 7.6, 11.7 and 21.9 percent more surveys completed in the High than Low Wage conditions for each of the quality restrictions. The interaction estimates for High Wage by Day 2 (Columns 4, 6 and 8) are all negative, suggesting that the high wage effect on surveys completed may be wearing off, yet none reach the level of significance ($p > .20$). Thus, the effect on quality adjusted output for paying the higher wage was significant and did not decrease significantly from Day 1 to Day 2.²⁶

The estimates for the effect of the Fair Framing are also positive for all levels of quality (Columns 3-8), but never reach the level of significance. Thus, we cannot reject the hypothesis that framing wages as fair (in the manner we did in this study) has no effect on surveys completed. There is also no statistical evidence that the Fair manipulation had any differential effect between Day 1 and Day 2 since all the interaction estimates between Fair and Day 2 are insignificant (Columns 4, 6 and 8).

Result 4: Adjusting for quality, the number of surveys completed is significantly higher when paying the higher wage, and is between 8 and 22 percent higher depending on the quality restriction. However, the Fair Framing did not significantly affect the number of surveys completed regardless of the quality restriction.

3.6 Is it Profitable to Pay the Higher Wage?

We have thus far examined whether workers reciprocate higher wages with a greater quantity and/or quality of output. While this reciprocity is necessary to justify the gift exchange hypothesis, it is not sufficient unless the value of the reciprocity is greater than the higher costs associated with paying the higher wage. To our knowledge, there is no field evidence showing that offering a higher fixed wage as a “gift” results in an increase in profits to justify the higher wage. For instance, even when workers were significantly more productive when paid the higher wage in the field experiments discussed above, the greater quantity of output was insufficient to offset the higher wages (footnote 5 discusses the evidence). If these results are robust, then the evidence suggests that the gift exchange hypothesis does not provide a plausible explanation for labor market equilibria in which profit maximizing firms pay wages above the market clearing wage. However, it is possible that offering a higher wage increased the unmeasured quality sufficiently to justify the higher wage.

Our results showed that paying 29 percent higher wages (\$22 rather than \$17) significantly improved the quality of the output and the quality adjusted quantity. However, it is not obvious from these results whether the improvement was sufficient to justify the higher wage. On first glance, it may seem that the improvement was insufficient since the increase in all the quality adjusted levels of output (which was between 8 and 22 percent) was less than the 29 percent higher wage. However, we should be skeptical of this simple conclusion since we assume that firms attach a cost to each error (or possibly multiple costs for

²⁶ Although the interaction High Wage by Day 2 is not significant, we further tested whether the effect of the High Wage had a significant effect when we analysed the data for Day 2 alone. For the 5%, 2% and 0% restrictions the p-values on High Wage are $p=0.13$, $p=0.07$ and $p=0.02$, respectively. Thus, for the 2% and 0% quality adjusted restrictions the High Wage had a significant effect on quality adjusted quantity on the second day.

different types of errors), implying that a one percent reduction in errors could be worth more (or less) than a one percent increase in costs. Thus, to determine the optimal wage, \$17 or \$22, the firm would compare the profits at each wage including an evaluation at the firm's marginal value per error.

The following analysis uses the experimental data to compute and compare the profitability for a range of marginal values of errors v , to determine under which values of v , if any, it would have been profitable to pay the higher "gift" wage. Specifically, we observe in each wage condition w ($w=\$17$ or $w=\$22$) that N_w total surveys were entered over all workers S_w , and E_w total errors were made, where $E_w = N_w * 51$ (radio button responses) * e_w , where e_w is the percent of errors across all data entry in w . If the firm has a constant marginal value v on the loss due to each error, then the number of surveys per dollar (Y_w) the firm obtains including the cost of errors is:

$$(3) \quad Y_w = N_w / [(w * S_w * (200/60)) + E_w * v],$$

where 200/60 is the number of hours each worker worked. From the data, we have $N_{17} = 1,825$, $S_{17} = 32$, $E_{17} = 1,962$, $N_{22} = 1,815$, $S_{22} = 31$ and $E_{22} = 1,533$. Thus, we have:

$$(4) \quad Y_{17} = 1,825 / [1813.33 + 1,962v] \text{ and } Y_{22} = 1,815 / [2,273.33 + 1,533*v].$$

Figure 7 shows the number of surveys per dollar for each value of v . It shows that when the marginal value of errors v is small, paying the \$17 wage is more profitable than the \$22 wage. Intuitively, this result follows since a similar number of surveys per worker are entered for a lower cost when there is a small penalty associated with errors. However, as the firm places an increasingly higher value on the errors, the higher wage condition becomes relatively more profitable as the higher errors in the low wage condition impose increasingly greater costs to the firm.²⁷ From (4) and as seen in Figure 7, if $v < \$0.89$ the firm maximizes profits by paying \$17 per hour ($Y_{17} > Y_{22}$), otherwise it maximizes profits paying \$22 per hour.²⁸ Thus, paying the higher "gift exchange" wage will be profitable if the firm places a sufficiently high value on errors, otherwise the gift exchange wages are not profitable.

Result 5: If the cost a firm places on errors is sufficiently high, in the current data entry above \$0.89, then it would be profitable to offer the gift exchange wage.

3.7 The Value of the Laboratory Controls

One advantage of having the real work done in the lab is the additional controls on the variation in the difficulty of the work (e.g., the packets) and the ability to measure individual differences in the workers' ability. By reducing other major sources of variation in productivity, we may be able to estimate "effort" more accurately, and thus detect treatment effects that are more difficult to measure outside the laboratory.

²⁷ In this analysis, we do not compute the firm's profit function. Instead, we assume that if the firm chooses the wage to minimize the cost to produce the identical number of surveys, this will maximize profits. However, it is potentially possible that there could be complementarities among workers (if workers could interact) or other alternative work conditions, such as providing bonus', imposing quality control requirements or using double entry procedures, would naturally affect the profit maximizing wage. The current analysis is thus for the current experimental conditions only.

²⁸ Note that the cost per survey in the \$17 and \$22 wage conditions is \$0.99 and \$1.25 and the error rate per survey is 1.08 and 0.84 radio buttons, respectively. Thus, the firm faces a trade off of 26% higher costs for a 22% reduction in errors, or about 1.18 percent higher costs to reduce errors by one percent.

The first two rows of Table 6 show the p-value for the High Wage treatment and the goodness of fit (R^2) from the models presented in Table 5 Columns 1, 3, 5 and 7 (the regressions without interaction terms). The last two rows show the p-values and R^2 from identical regressions omitting the controls for packet, average characters and maximum ability.²⁹ Using conventional significance levels, i.e., $p < .05$ for significance and $.05 < p < .10$ for marginal significance, we find significant effects of the High Wage treatment for the three quality thresholds when we include the controls. However, without these additional controls, we would have only detected a significant High Wage treatment effect for the strictest quality threshold and a marginally significant effect for the 2% threshold. The insert also shows that the model fit improves substantially with the inclusion of the controls. For instance, R^2 increases by a factor of between almost 4 (with No Errors) to a factor of over 6 (with no quality threshold imposed).

Table 6: Significance of High Wage Treatment Effect

| | | Quality Standards | | | |
|-------------------------|----------------------|---------------------|-----------------------|-----------------------|------------------|
| | | <i>None Imposed</i> | <i>< 5% Errors</i> | <i>< 2% Errors</i> | <i>No Errors</i> |
| with controls | p-value | 0.263 | 0.036 | 0.014 | 0.007 |
| | R² | 0.600 | 0.420 | 0.332 | 0.249 |
| without controls | p-value | 0.566 | 0.192 | 0.062 | 0.030 |
| | R² | 0.087 | 0.079 | 0.080 | 0.063 |

Result 6: The additional controls provided the necessary precision to detect significant effects of the High Wage for all three quality thresholds.

4. Summary

We conducted a labor market experiment in the laboratory. Workers (subjects) were recruited to enter data for a large non-profit organization over two days one week apart. The experiment introduces a hybrid methodology that combines desirable features of laboratory and field experiments. Although physical and mental effort is not new to the lab (many experiments use real effort tasks such as adding numbers and solving mazes), the current method goes beyond physical and mental effort; workers produced output that remains valuable outside of the experimental context and they were paid by the firm. This closer distance to a job outside the lab was used to capture qualitative aspects of work that may differ between the laboratory and the field, thus improving the external validity of the experiment, while the laboratory let us collect additional controls (e.g., difficulty of the work with the packets) and measures of individual differences (e.g., maximum ability) that may be difficult to observe in a natural setting. The hybrid method thus gives us greater ability to observe and control all dimensions of the work, some of which may be substantially more costly or noisy to measure in a more natural setting. To the extent that there may be many other hard to observe dimensions to the work (e.g., helping others, complementarities across workers, many types of errors or other quality dimensions), our design essentially eliminates these alternative avenues to reciprocate. As we have shown, this is critical to effectively test the gift-exchange

²⁹ The R^2 are adjusted for the number of parameters estimated in each model.

hypothesis since both previous empirical work and theory points toward difficult to observe dimensions of work as potentially being the relevant margin in which workers may reciprocate higher wages.

We find that workers in the High Wage treatment perceived their wages to be significantly fairer than workers in the Low Wage condition and workers in the Fair Framing also perceived their wages to be fairer than workers in the Unfair Framing. This provides direct evidence that both the High Wage and the Fair Framing led to a higher perception of the fairness of the wage.

Our results show that, consistent with some of the past field studies, paying a higher wage did not significantly increase the total number of surveys completed when quality is ignored. We draw the same conclusion even after controlling for the variability in the work and heterogeneity in the workers ability. However, including the controls increased the precision of the estimates of the treatment effects substantially; the standard errors fell by nearly 50 percent and the model fit increased dramatically. We also find that while workers being paid the higher wage did not enter more surveys overall, they made significantly fewer errors. The workers in the higher wage treatment also produced more surveys once we control for the quality of the output.³⁰ These results support the importance of examining the multi-dimensional nature of effort when measuring how workers can reciprocate gift exchange higher wages. We thus find that reciprocal preferences occur and persist over two days of work one week apart in a work environment involving a real organization requiring meaningful output.

While workers receiving higher wages made significantly fewer errors and entered significantly more surveys when we imposed quality restrictions, we never detect a significant difference in quality or the number of surveys entered (overall or adjusted for quality) between the Fair and Unfair framing of the wages. This is interesting given that the perception of the fairness of the wage (measured before and after doing the work) indicates that the fair framing had a virtually identical affect on perceptions as the wage manipulation. This suggests that the reason the higher wage increased quality and quality adjusted output could be due to some other or additional psychological channel than fairness perceptions.³¹ For instance, Kube et al (2011) find worker productivity increased with an in-kind gift and Kosfeld and Neckermann (2011) find that a purely symbolic gift also increased productivity. These results suggest that employer effort to offer a gift may also be a critical channel for the gift exchange.

Unlike prior field experiments studying the gift exchange hypothesis, we also find that it can be profitable for firms to pay a higher fixed wage due to reciprocal worker preferences. In our data entry work, if the firm places a monetary value of \$0.89 or more on each error, then it would have been

³⁰ This is *not* at odds with the past field-studies; For instance, List (2008) notes that the number of doors knocked in a charity solicitation task (Gneezy and List (2006) is indistinguishable by treatment, only the *amounts* raised by each treatment differ. This again points toward the importance of distinguishing all relevant dimensions of work when testing for reciprocal behaviour.

³¹ One possible explanation, following Englmaier and Leider (2011), is that the response to the wage is driven by reciprocal concerns, with perceived fairness being only one component of the utility gained from reciprocation. Since the work was being conducted for a large pro-social organisation, concerns for reciprocation to the organization may have had a comparatively minor effect on their behaviour. It may also have been that the question with which fairness perceptions were gathered (“*Although we are paying you \$17(\$22) per hour, we would like to know how fair you would think other possible wages might have been. For each wage listed, please click on the response, from Extremely Unfair (1) to Extremely Fair (7) that best reflects how you would feel about getting paid each wage for the data entry task you will be doing today and next week.*”) did not accurately reflect workers feelings of fairness toward the firm, but rather a more general account of fairness.

profitable to pay the higher wage. Inducing reciprocal behavior with higher wages may thus be most profitable when the value of quality is high enough and more traditional methods of motivating workers is not feasible – either due to the nature of the task, or the costs associated with monitoring the quality of the work.

The current and previous field experiments examine data entry with only a single worker entering each observation. However, many firms use double entry (two workers enter the same data); almost every firm we found online advertises using double entry.³² Thus, an extension to the current work may examine what value of errors, if any, would have been profitable to pay the higher wage with double entry. We analyzed the current data under many assumptions³³ on how data entry might differ in single vs. double entry and we always find that the results for double entry are qualitatively similar to single entry; intuitively, since double entry reduces the number of uncorrected errors roughly proportionally for each wage condition after identified mismatches are corrected, under double entry there will be fewer uncorrected errors in the higher wage condition, and so the higher cost the firm places on these uncorrected errors, the relatively more profitable the higher wage.

In sum, we find worker behavior consistent with reciprocal preferences; workers receiving higher wages produced significantly higher effort that persisted over two days of work one week apart. The greater effort occurred on the harder to observe quality dimension and was not detected on the easily observable quantity dimension. Moreover, depending on the value placed on the harder to observe quality dimension of the work, it would have been profitable to pay the higher wage. These results imply that the extent to which firms are able to monitor workers effectively, and the value of the qualitative elements which go undetected, may be the critical factors that gives relevance to the “gift-exchange” hypothesis in labor market behavior.

³² Five major data entry firms that indicate double entry. <http://www.dataentryservices.co.in/>; <http://www.axiondata.com/>; <http://www.owdata.com/>; <http://www.relyservices.com/>; <http://www.dataentrysolution.com/>.

³³ There appear to be four major assumptions in order to use single entry data to extrapolate to double entry. First, we need to assume that workers in double entry would have been equally productive in both quantity and quality as they were in single entry, or at least proportionally equivalent changes would have occurred using double rather than single entry across the two wages. Second, we need to make an assumption about the marginal costs to identify errors (this may arguably be nothing with conventional software). Third, we need to make assumptions about the distribution of the coding errors. Whether errors are made systematically or randomly can affect the number of errors that would be detected and undetected in double entry. Fourth, we need to make assumptions regarding the costs and likelihood of successfully correcting mismatched entries.

References

- Akerlof, George A.** "Labor Contracts as Partial Gift Exchange." *The Quarterly Journal of Economics*, 1982, 97(4), 543-569.
- Al-Ubaydli, Omar; Andersen, Steffen; Gneezy, Uri and List, John A.** "For Love or Money? Comparing the Effects of Non-Pecuniary and Pecuniary Incentive Schemes in the Workplace," working paper, 2008.
- Baker, George; Gibbons, Robert and Murphy, Kevin J.** "Subjective Performance Measures in Optimal Incentive Contracts." *The Quarterly Journal of Economics*, 1994, 109, 1125-1156.
- Charness, Gary; Frechette, Guillaume R. and Kagel, John H.** "How Robust Is Laboratory Gift Exchange?" *Experimental Economics*, 2004, 7(2), 189-205.
- Cohn, Alain; Goette, Lorenz and Fehr, Ernst.** "Fairness and Effort," working paper, 2009.
- Englmaier, Florian and Leider, Stephen.** "Gift Exchange in the Lab – it is not (only how much you pay ...)," Working paper, 2009.
- Englmaier, Florian and Leider, Stephen.** "Contractual and Organizational Structure with Reciprocal Agents," *American Economic Journal – Microeconomics*, 2011 (forthcoming).
- Fehr, Ernst; Kirchsteiger, Georg and Riedl, Arno.** "Does Fairness Prevent Market Clearing? An Experimental Investigation." *The Quarterly Journal of Economics*, 1993, 108(2), 437-459
- Fehr, E. and K. M. Schmidt.** "Fairness and Incentives in a Multi-Task Principal–Agent Model." *Scandinavian Journal of Economics*, 2004, 106(3), pp. 453-74..
- Gneezy, Uri and List, John A.** "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments." *Econometrica*, 2006, 74(5), 1365-1384.
- Hennig-Schmidt, H., Rockenbach, B. and Sadrieh, A.** "In Search of Workers' Real Effort Reciprocity--A Field and a Laboratory Experiment," *Journal of the European Economic Association*, 2010, 8, 817-837.
- Holmstrom, Bengt and Milgrom, Paul.** "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization*, 1991, 7(Papers from the Conference on the New Science of Organization, January 1991), 24-52.
- Kapoor, Sacha** "Incentive Provision in Multitask Jobs: Experimental Evidence from the Workplace", Working Paper 2010.
- Kosfeld, Michael and Neckermann, Suzanne** "Getting More Work for Nothing? Symbolic Awards and Worker Performance", *American Economic Journal: Microeconomics*, 2011, forthcoming.
- Krueger, Alan B. and Mas, Alexandre.** "Strikes, Scabs, and Tread Separations: Labor Strife and the Production of Defective Bridgestone/Firestone Tires." *The Journal of Political Economy*, 2004, 112(2), 253-289.
- Kube, Sebastian; Maréchal, Michel André and Puppe, Clemens.** "Putting Reciprocity to Work - Positive Versus Negative Responses in the Field," working paper, 2008.
- Kube, Sebastian; Maréchal, Michel André and Puppe, Clemens.** "The Currency of Reciprocity---Gift Exchange in the Workplace," *American Economic Review*, 2011, forthcoming.
- Levitt, Steven D. and List, John A.** "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *The Journal of Economic Perspectives*, 2007, 21(2), 153-174.
- List, John A.** "Social Preferences: Some Thoughts from the Field." *Annual Review of Economics*, 2009, 1(1), 563-583.
- Mas, Alexandre.** "Pay, Reference Points, and Police Performance*." *Quarterly Journal of Economics*, 2006, 121(3), 783-821.

- Mas, Alexandre and Moretti, Enrico.** "Peers at Work." *American Economic Review*, 2009, 99(1), 112-145.
- Prendergast, Canice.** "The Provision of Incentives in Firms." *Journal of Economic Literature*, 1999, 37(1), pp. 7-63.
- Slonim, Robert, and Garbarino, Ellen.** "The Effect of Price History on Demand as Mediated by Perceived Price Expensiveness," *Journal of Business Research*, 1999, 45, 1-14.
- Xiao, Erte and Houser, Daniel.** "Emotion Expression in Human Punishment Behavior." *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(20), 7398-7401

Table 1: Participant Summary Statistics

| | <u>Unfair \$17</u> | <u>Fair \$17</u> | <u>Unfair \$22</u> | <u>Fair \$22</u> | <u>All</u> |
|-----------------------------------|--------------------|------------------|--------------------|------------------|-----------------|
| Participants | 16 | 16 | 15 ^a | 16 | 63 |
| Women | 7 | 5 | 8 | 7 | 27 |
| Mean Age | 21.2 | 22.4 | 21.1 | 22.2 | 21.7 |
| Had Experience With Data-Entry | 5 | 9 | 6 | 7 | 27 |
| Currently Employed | 6 | 8 | 7 | 10 | 31 |
| Economics or Business Major | 7 | 6 | 6 | 7 | 26 |
| Doubted Experiment | 2 | 1 | 0 | 3 | 6 |
| Mean Piece-Rate Payment (s.d.) | 22.53 (3.42) | 23.49 (4.61) | 24.08 (3.56) | 22.19 (4.22) | 23.06 (3.96) |

^a One participant arrived over 60 minutes late on Day 2 and was excluded from participating that day. We exclude this participant from all of the analyses.

Table 2: Manipulation Check*Dependent Variable: Fairness of Wage.^a 1=Extremely Unfair; 7=Extremely Fair*

| | Tobit Regressions | | | Ordered Probit Regressions | | | |
|-------------------------|---------------------|---------------------|---------------------|----------------------------|---------------------|---------------------|-------|
| | Day 1 | Day 2 | Both Days | Day 1 | Day 2 | Both Days | |
| Constant | 3.572*** (0.395) | 2.866*** (0.309) | 3.571*** (0.352) | | | | |
| Fair Framing | 1.896*** (0.479) | 1.705*** (0.371) | 1.862*** (0.426) | 1.109*** (0.295) | 1.233*** (0.288) | 1.179*** (0.285) | |
| High Wage | 1.965*** (0.486) | 2.359*** (0.373) | 1.913*** (0.413) | 1.144*** (0.294) | 1.689*** (0.308) | 1.200*** (0.284) | |
| Day 2 | | | -0.740* (0.427) | | | -0.481 (0.324) | |
| Fair Framing x Day 2 | | | -0.0864 (0.516) | | | -0.0641 (0.386) | |
| High Wage x Day 2 | | | 0.521 (0.520) | | | 0.357 (0.389) | |
| Log-Likelihood | -101.78 | -97.42 | -198.50 | Pseudo R ² | 0.135 | 0.194 | 0.162 |
| N | 62 | 62 | 124 | | 62 | 62 | 124 |

* p<.10; ** p<.05; *** p<.01.

^a The survey question was: "Although we are paying you \$17(\$22) per hour, we would like to know how fair you would think other possible wages might have been. For each wage listed, please click on the response, from Extremely Unfair (1) to Extremely Fair (7) that best reflects how you would feel about getting paid each wage for the data entry task you will be doing today and next week." We dependent variable is the response to the wage subjects actually received, thus the responses for participants in Fair and Unfair \$17 is for the \$17 wage and for participants in Fair and Unfair \$22 is for the \$17 wage.

Table 3: Quantity Produced Unadjusted for Quality

Dependent Variable: Surveys completed per hour

OLS Estimates (Robust Standard Errors)

| | (1) | (2) | (3) | (4) | (5) |
|-----------------------------|---------------------|---------------------|----------------------|----------------------|----------------------|
| Constant | 17.21*** (0.651) | 13.07*** (1.611) | 14.80*** (1.579) | -0.387 (1.718) | -0.342 (1.739) |
| Fair Framing | 0.236 (0.854) | 0.221 (0.743) | 0.316 (0.761) | 0.635 (0.429) | 0.423 (0.469) |
| High Wage | 0.497 (0.860) | 0.697 (0.739) | 0.668 (0.762) | 0.529 (0.473) | 0.771 (0.494) |
| Average Characters | | | -0.033*** (0.002) | -0.033*** (0.002) | -0.033*** (0.002) |
| Piece Rate Output | | | | 0.784*** (0.068) | 0.783*** (0.068) |
| Fair Framing x Day 2 | | | | | 0.427 (0.458) |
| High Wage x Day 2 | | | | | -0.474 (0.499) |
| <i>R</i> ² | 0.003 | 0.235 | 0.334 | 0.600 | 0.601 |
| Observations | 1134 | 1134 | 1134 | 1134 | 1134 |
| Participants | 63 | 63 | 63 | 63 | 63 |
| Errors Clustered by Worker | Yes | Yes | Yes | Yes | Yes |
| Control for Time & Packet | No | Yes | Yes | Yes | Yes |
| Other Controls ^a | No | Yes | Yes | Yes | Yes |

* p<.10; ** p<.05; *** p<.01

^a Other controls: (1) Doubt (on Day 2 Survey): "Did you doubt any aspect of the experiment? yes or no, (2) Know organization (on Day 1 Survey): "Prior to coming here today, how well did you know the [charitable organization]? From 1=Not at All to 7=Extremely Well?, (3) Tiredness (on Day 1 Survey) "The following statements are about how you are feeling right now. For each statement, please click on the response, from Strongly Disagree (1) to Strongly Agree (7) that best reflects how you are feeling: I feel tired," Tiredness (on Day 2 Survey): "Now that you have completed the one hour and forty minutes of data entry work for today, how tired do you feel right now, on a scale from 1 (not at all tired) to 7 (very tired)?"

Table 4: Quality Regressions
 Dependent Variable: Percent Errors
 Tobit Estimates (Robust Standard Errors)

| | (1) | (2) | (3) | (4) | (5) |
|----------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Constant | 2.201*** (0.239) | 3.954*** (0.593) | 3.944*** (0.598) | 2.510*** (0.898) | 2.669*** (0.908) |
| Fair Framing | 0.085 (0.278) | 0.094 (0.224) | 0.094 (0.224) | 0.121 (0.216) | -0.172 (0.267) |
| High Wage | -0.591** (0.278) | -0.545** (0.231) | -0.545** (0.231) | -0.563** (0.224) | -0.442* (0.267) |
| Average Characters | | | 0.000 (0.001) | 0.000 (0.001) | 0.000 (0.001) |
| Piece Rate Output | | | | 0.072** (0.0343) | 0.071** (0.0344) |
| Fair Framing x Day 2 | | | | | 0.592** (0.292) |
| High Wage x Day 2 | | | | | -0.231 (0.290) |
| Log-Likelihood | -2566.15 | -2536.12 | -2536.11 | -2533.99 | -2531.65 |
| Observations | 1134 | 1134 | 1134 | 1134 | 1134 |
| Participants | 63 | 63 | 63 | 63 | 63 |
| Errors Clustered by Worker | Yes | Yes | Yes | Yes | Yes |
| Control for Time & Packet | No | Yes | Yes | Yes | Yes |
| Other Controls | No | Yes | Yes | Yes | Yes |

* p<.10; ** p<.05; *** p<.01

^a Other controls: (1) Doubt (on Day 2 Survey): “Did you doubt any aspect of the experiment? yes or no, (2) Know organization (on Day 1 Survey): “Prior to coming here today, how well did you know the [charitable organization]? From 1=Not at All to 7=Extremely Well?, (3) Tiredness (on Day 1 Survey) “The following statements are about how you are feeling right now. For each statement, please click on the response, from Strongly Disagree (1) to Strongly Agree (7) that best reflects how you are feeling: I feel tired,” Tiredness (on Day 2 Survey): “Now that you have completed the one hour and forty minutes of data entry work for today, how tired do you feel right now, on a scale from 1 (not at all tired) to 7 (very tired)?

Table 5: Quantity of Surveys with Quality Restrictions
 Dependent Variable: Surveys completed per hour that meet quality standards
 OLS Estimates (Robust Standard Errors)

| | Quality Standards | | | | | | | |
|---------------------------|------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| | None Imposed | | < 5% Errors | | < 2% Errors | | No Errors | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Constant | -0.387 (1.718) | -0.342 (1.739) | -1.430 (2.253) | -1.604 (2.280) | -1.001 (2.602) | -1.486 (2.653) | -1.059 (2.304) | -1.426 (2.270) |
| Fair Framing | 0.635 (0.429) | 0.423 (0.469) | 0.593 (0.489) | 0.723 (0.579) | 0.849 (0.610) | 1.177 (0.779) | 0.365 (0.542) | 0.605 (0.780) |
| High Wage | 0.529 (0.473) | 0.771 (0.494) | 1.127** (0.537) | 1.261** (0.601) | 1.557** (0.632) | 1.980** (0.783) | 1.538*** (0.571) | 1.872** (0.825) |
| Average Characters | - 0.033** * (0.002) | -0.033*** (0.002) | -0.033*** (0.003) | -0.033*** (0.003) | -0.030*** (0.003) | -0.029*** (0.003) | -0.018*** (0.003) | -0.018*** (0.003) |
| Piece Rate Output | 0.784** * (0.068) | 0.783*** (0.068) | 0.676*** (0.079) | 0.676*** (0.079) | 0.555*** (0.098) | 0.557*** (0.089) | 0.261*** (0.089) | 0.262*** (0.089) |
| Fair Framing X Day 2 | | 0.427 (0.458) | | -0.263 (0.779) | | -0.663 (0.959) | | -0.486 (1.103) |
| High Wage x Day 2 | | -0.474 (0.499) | | -0.274 (0.884) | | -0.861 (1.036) | | -0.681 (1.116) |
| R² | 0.5999 | 0.6013 | 0.4196 | 0.4198 | 0.3318 | 0.3332 | 0.2486 | 0.2494 |
| Observations | 1134 | 1134 | 1134 | 1134 | 1134 | 1134 | 1134 | 1134 |
| Participants | 63 | 63 | 63 | 63 | 63 | 63 | 63 | 63 |
| Acceptable Surveys | 3671 | 3671 | 3270 | 3270 | 2835 | 2835 | 1486 | 1486 |
| Errors Clusted by Worker | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Control for Time & Packet | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Other Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |

* p<.10; ** p<.05; *** p<.01

^a Other controls: (1) Doubt (on Day 2 Survey): “Did you doubt any aspect of the experiment? yes or no, (2) Know organization (on Day 1 Survey): “Prior to coming here today, how well did you know the [charitable organization]? From 1=Not at All to 7=Extremely Well?, (3) Tiredness (on Day 1 Survey) “The following statements are about how you are feeling right now. For each statement, please click on the response, from Strongly Disagree (1) to Strongly Agree (7) that best reflects how you are feeling: I feel tired,” Tiredness (on Day 2 Survey): “Now that you have completed the one hour and forty minutes of data entry work for today, how tired do you feel right now, on a scale from 1 (not at all tired) to 7 (very tired)?

Figure 1: Distribution of Payment in the Piece Rate Session

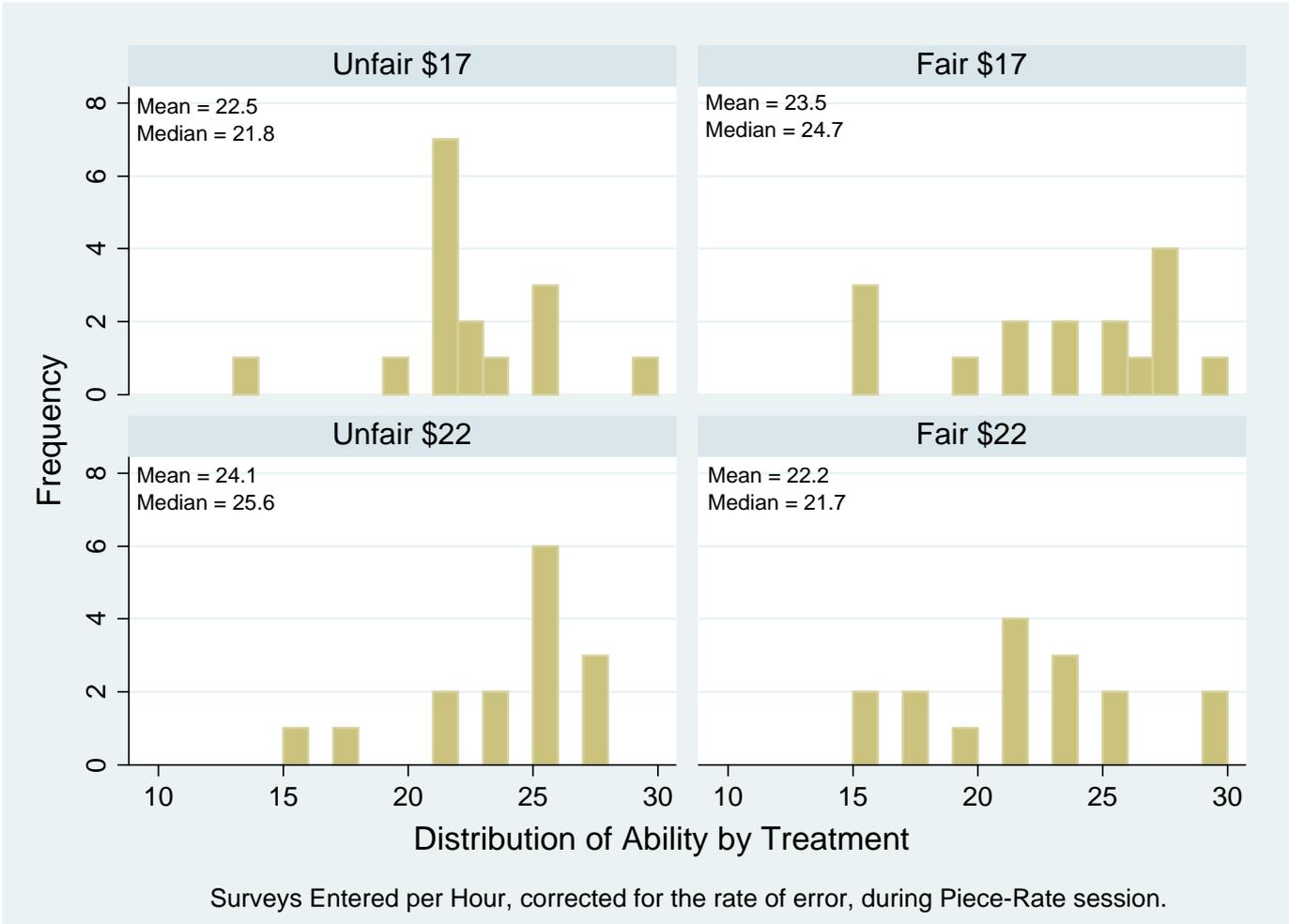


Figure 2a: Fairness of Wage Perceptions on Day 1



Figure 2b: Fairness of Wage Perceptions on Day 2

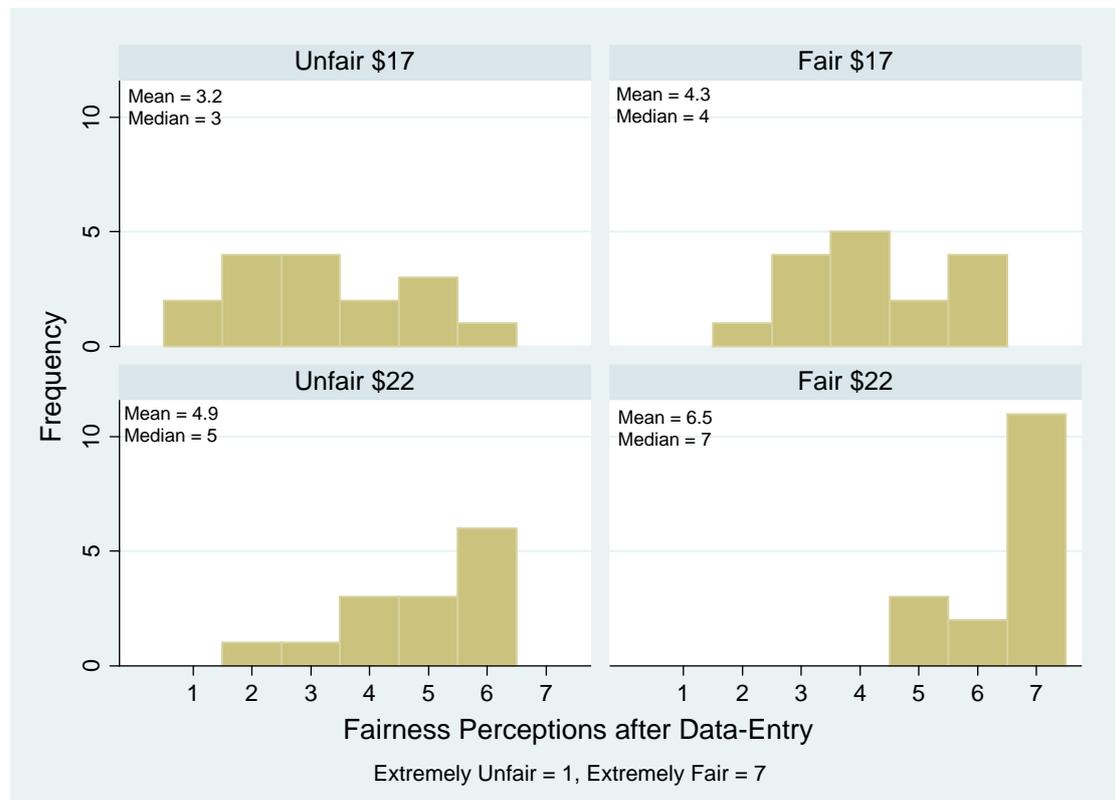


Figure 3a: All Surveys Completed over Time

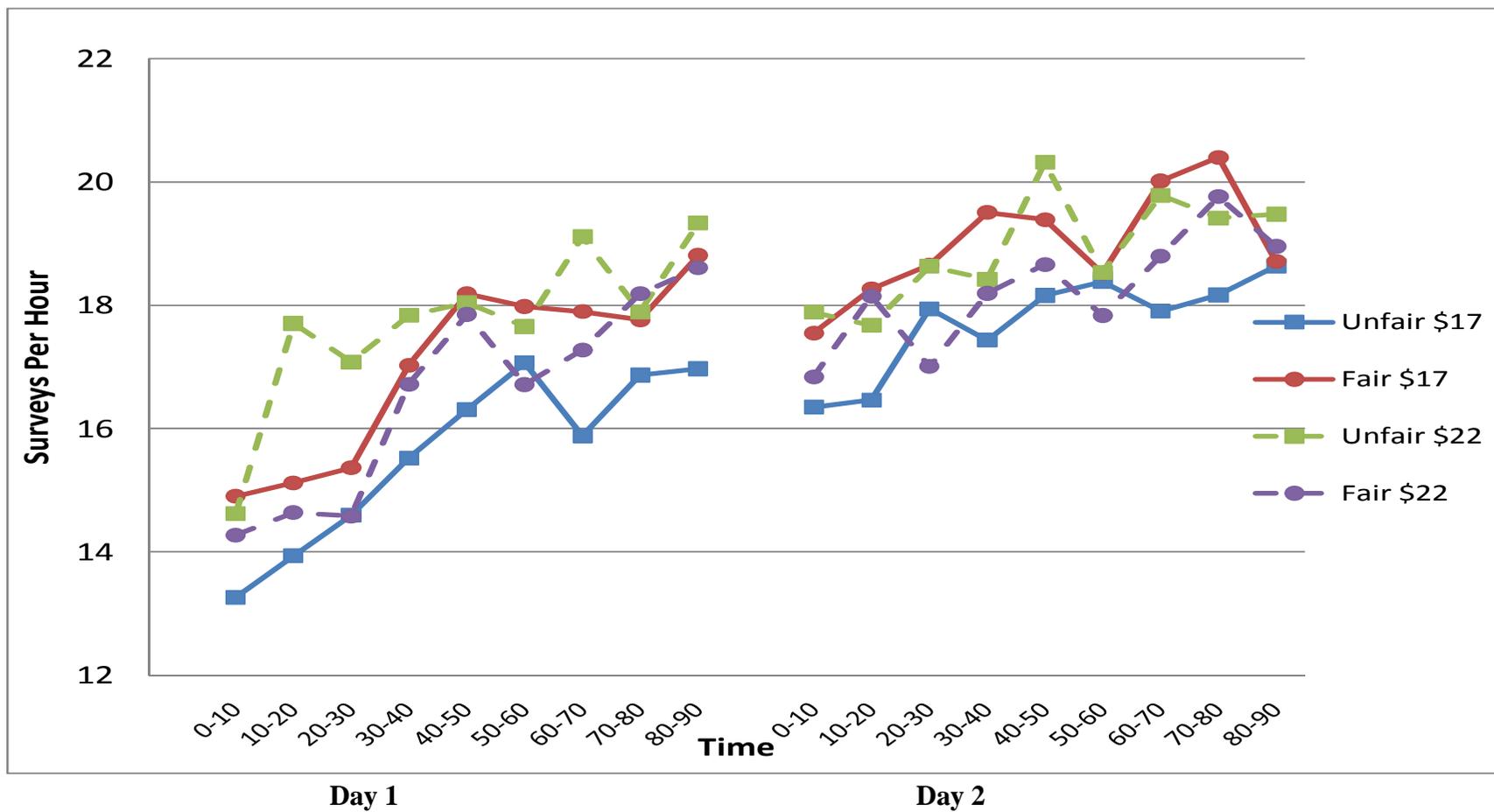


Figure 3b: Surveys Completed Time Meeting the 2% Quality Threshold

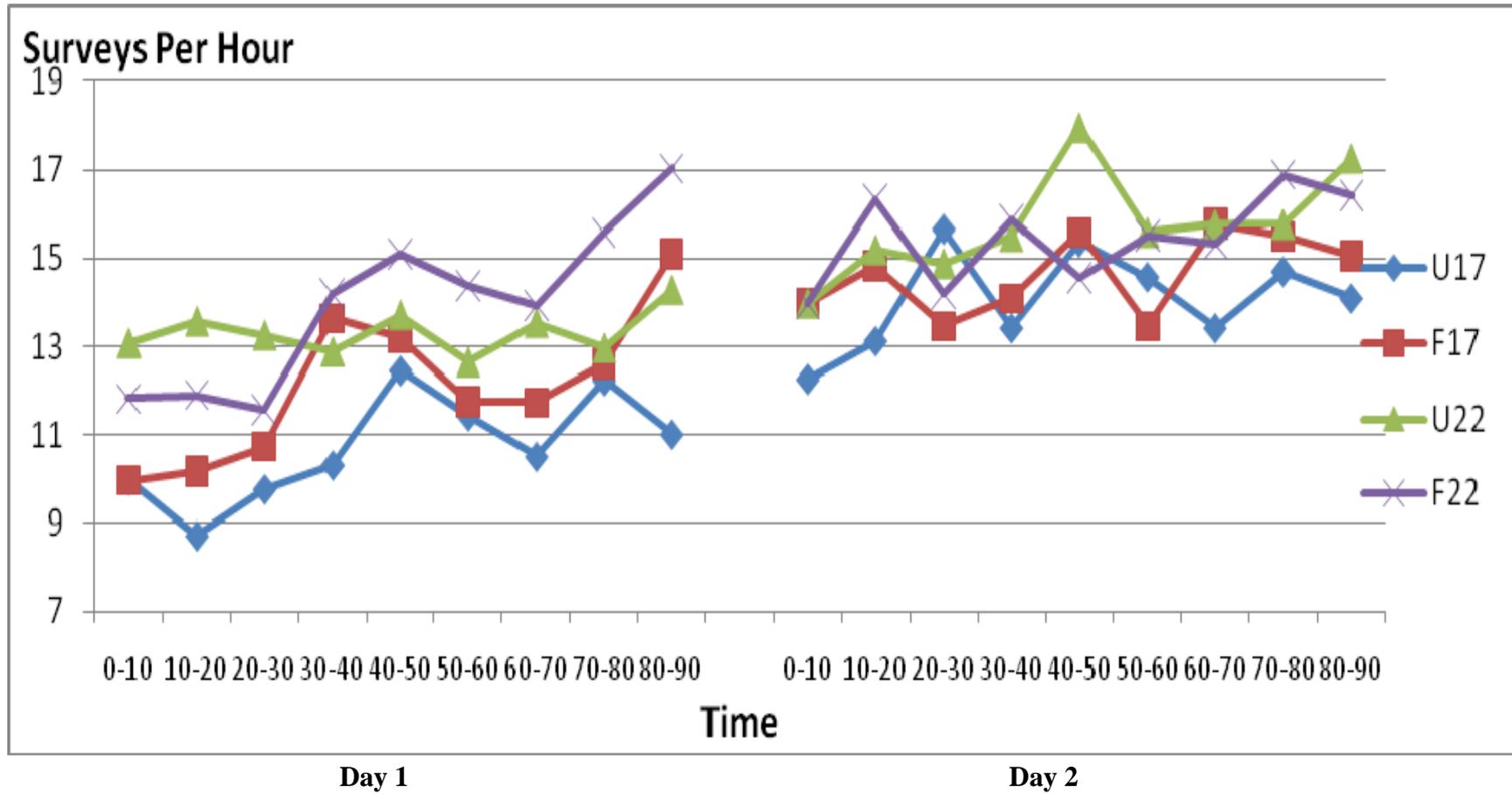


Figure 4: Average total surveys completed by Day and Treatment

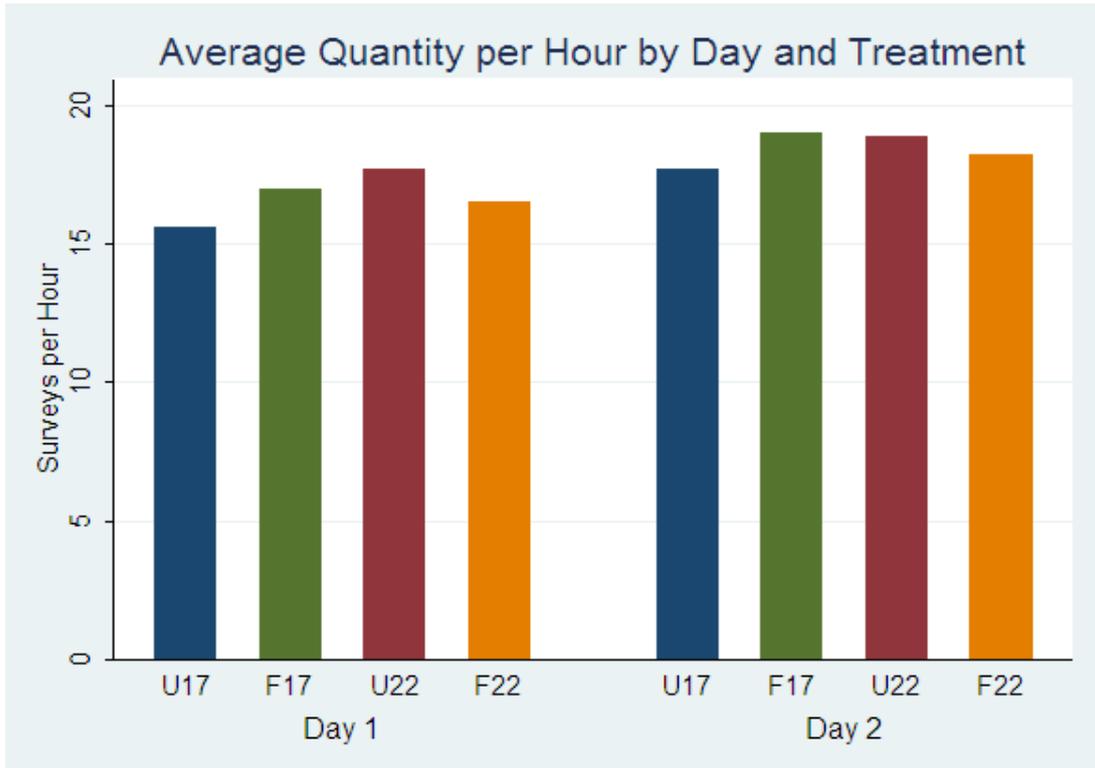
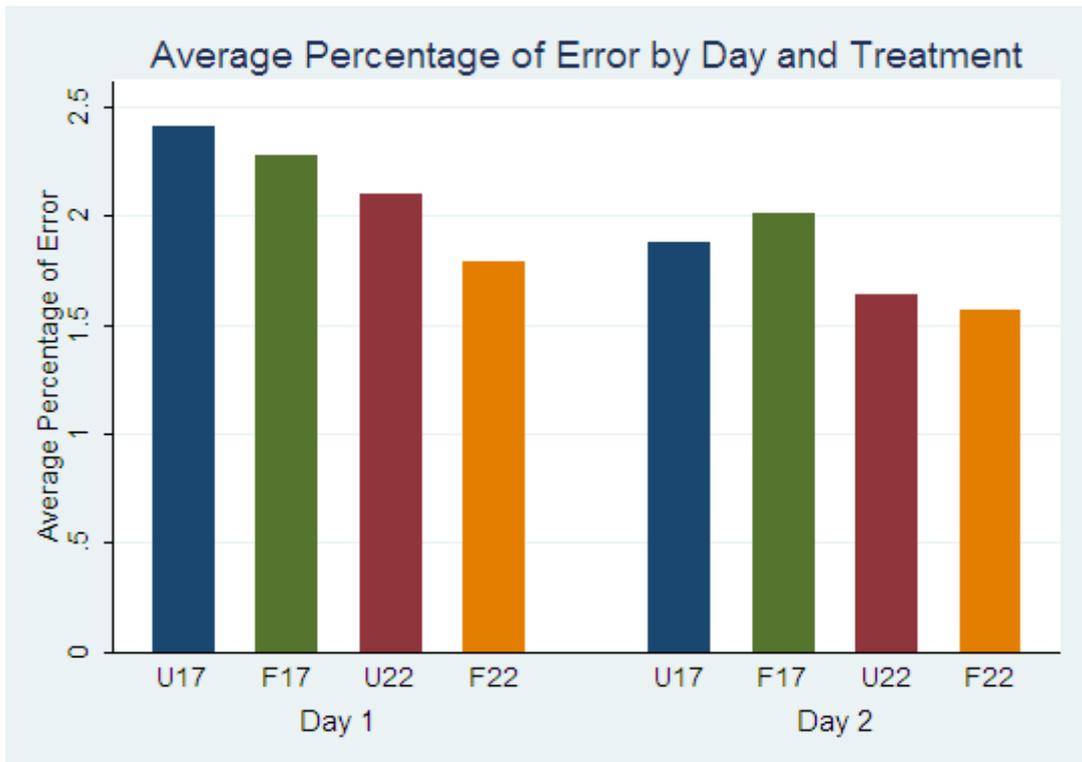


Figure 5: Average Percentage of Error by Treatment and Day
(One outlier Omitted)



The outlier observation occurs on Day 2 in the Fair \$17 treatment where one worker made 11.6 percent errors, nearly 7 times the average of everyone else on Day 2.

Figure 6a: Average Quantity per Hour of Acceptable Surveys by Treatment – Day 1

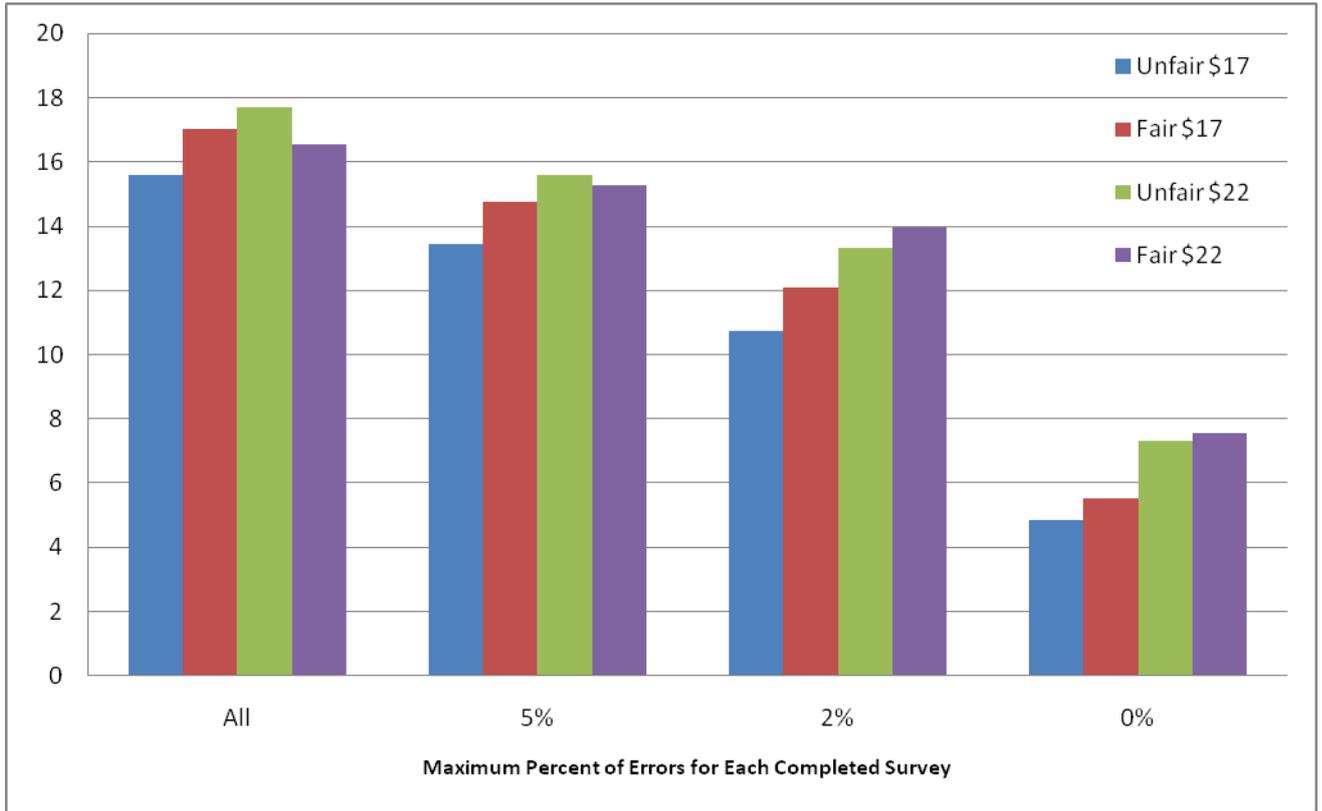


Figure 6b: Average Quantity per Hour of Acceptable Surveys by Treatment – Day 2

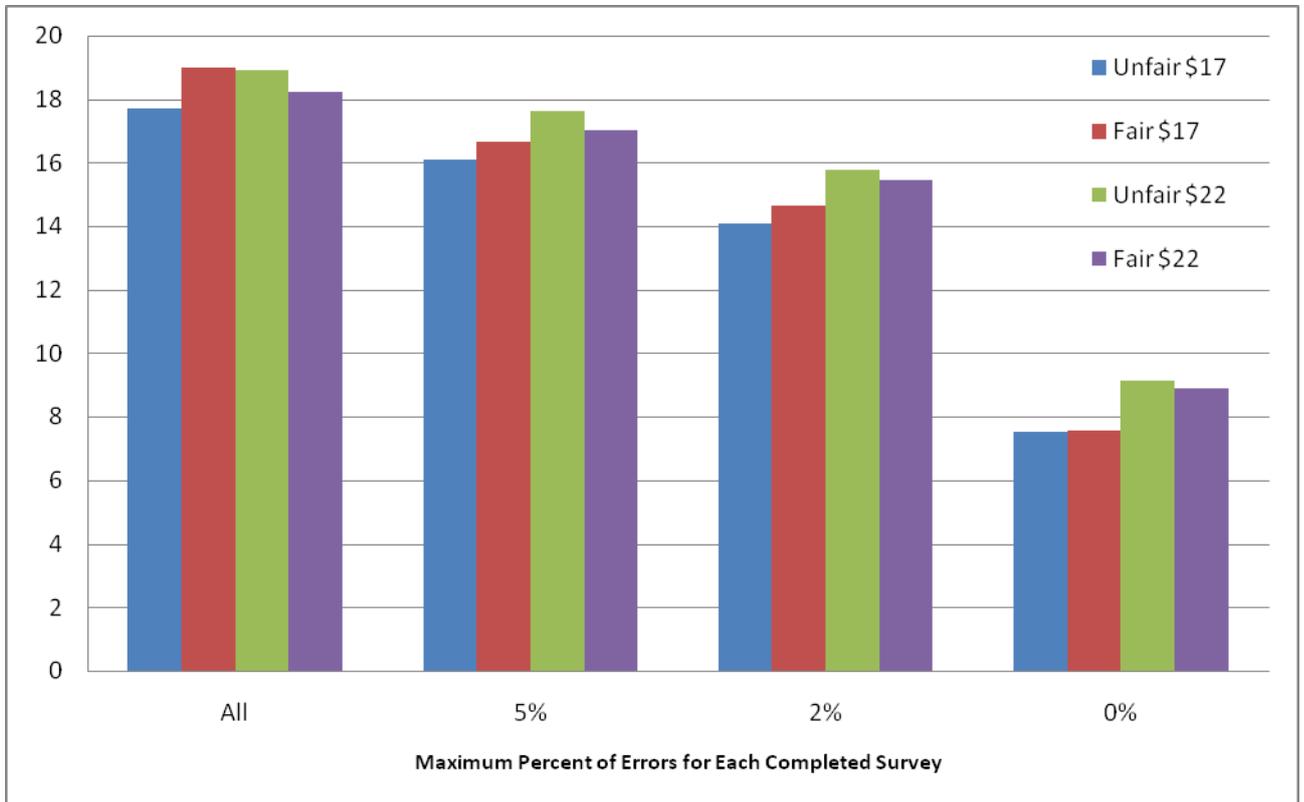


Figure 7: Number of Surveys per Dollar with Single Entry for \$17 and \$22 per hour wages

