Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017)

# When Does More Money Work? Examining the Role of Perceived Fairness in Pay on the Performance Quality of Crowdworkers

**Teng Ye, Sangseok You, Lionel P. Robert Jr.**
University of Michigan, Ann Arbor
{tengye, sangyou, lprobert}@umich.edu

**Abstract**

This research adds to the rich discussion on whether increases in payment to crowdworkers lead to increases in performance quality by introducing the concept of perceived fairness in pay (PFP). PFP refers to the belief that one is fairly compensated for their work. We examine whether PFP mediates the impact of payment amount on the performance quality of crowdworkers. We conducted a field experiment with 152 crowdworkers performing a button-clicking (BC) task and an instructional manipulation check (IMC) task. PFP mediated the impact of payment amount on performance quality in the BC task but not in the IMC task. PFP also mediated the impact of payment amount on satisfaction and task time. Results suggest that PFP can help us better understand the relationship between payment and performance quality in crowdsourcing.

## Introduction

Researchers have been interested in whether financial incentives can increase the performance of crowdworkers (Mason and Watts 2009; Yin and Chen 2016; Yin, Chen, and Sun 2013). Originally, it was assumed that increases in payment should lead to better performance. At the onset, the logic seemed intuitive: As payment increased so should the motivation for crowdworkers to work harder and hence perform better. However, the results have not been as straightforward.

Research seems to indicate that increases in payment can lead to increases in work performed (e.g., more tasks completed) but not performance quality (e.g., more tasks completed correctly; Buhrmester, Kwang, and Gosling 2011; Mason and Watts 2009; Rogstadius et al. 2011). However, recently, Ho et al. (2015) conducted a comprehensive study on payment and performance quality of crowdworkers showing the opposite results. They found that payment increases in the form of bonuses can increase performance quality but only for effort-responsive tasks. However, according to Ho et al. (2015) the bonuses must be big enough to be salient.

This research adds to this discussion in the following ways. First, we do not view crowdworkers as simply human central processing units (CPUs) that react rationally to the inputs given to them. Research on behavioral economics and psychology demonstrates that individuals can react differently to the same incentives (Folger 1977; Kahneman 2003; Tversky and Kahneman 1992). Their reactions often depend on the context, their perception of the incentives, their personal motivations, and their beliefs about their abilities (Camerer et al. 1997; Folger and Konovsky 1989; Greenberg and Colquitt 2013; Tversky and Kahneman 1992). Second, to better understand human workers' responses to incentives, we introduce the concept of perceived fairness in pay (PFP). PFP refers to the belief that one is being fairly compensated. PFP has been used in organizational studies to explain the relationship between financial incentives and an employee's performance, motivation, and satisfaction (e.g., Folger 1977; Folger and Konovsky 1989). We propose that in crowdsourcing, when crowdworkers believe they are being paid fairly they perform better. Third, we extend the conversation to include crowdworker satisfaction. Little research has been directed at understanding how or whether payment leads to more satisfied crowdworkers. Yet, as we envision a better future for crowd work (see Kittur et al. 2013), outcomes like satisfaction become increasingly important.

To do this, we conducted a field experiment involving 152 crowdworkers performing two types of tasks. One task was a button-clicking (BC) task, which represents an effort-responsive task, and the other an instructional manipulation check (IMC) task, which represents a non-effort-responsive attention-based task. Our results showed that PFP mediated the impact of the payment amount on the performance quality in the BC task but not the performance quality in the IMC task, satisfaction and even total task completion time.

This paper contributes to the ongoing conversation on the impact of financial incentives in crowd work in several important ways. First, it helps to answer the question of why increases in payment amount might or might not lead to better performance quality in effort-responsive tasks. When increases in payment amount lead to increases in PFP, they are likely to translate to better performance quality in effort-responsive tasks. Second, we also show that PFP can explain when increases in pay lead to more satisfied crowdworkers. Finally, the results of our study contribute to our understanding on how to better design crowd systems. Systems designed to help facilitate PFP can promote more effective and satisfactory crowd work.

## Related Work and Hypotheses Development

Scholars have been interested in the relationship between pay and performance in part to help understand how to minimize poor performance due to a lack of effort or attention on the part of crowdworkers (Goodman, Cryder, and Cheema 2013; Kittur et al. 2013). A complete review of this literature is beyond the scope of this paper. However, both Ho et al. (2015) and Yin et al. (2013) provide a thorough discussion on the topic. What is clear is that PFP has not been used to examine this problem.

Organizational scholars have also sought to understand when pay increases lead to better performance or more satisfied workers (Greenberg and Colquitt 2013), in large part because of similar concerns regarding poor performance by workers. These scholars have discovered that the actual amount paid is not necessarily related to an employee's performance or satisfaction (Jirjahn 2016; Tekleab, Bartol, and Liu 2005). Instead, changes in PFP explain why changes in payment can lead to increases in performance and satisfaction (Folger 1977; Folger and Konovsky 1989).

PFP is based on Adams' (1965) equity theory, which asserts that individuals attempt to match their effort with their associated outcomes. Individuals expect to be compensated relative to the amount of effort they believe is required to perform their job (Greenberg and Colquitt 2013; Saunders and Thornhill 2003). Next, we explain how this applies in the context of crowdworkers.

PFP should mediate the relationship between payment amount and performance quality of crowdworkers for several reasons. First, actual payment should be positively associated with perceived fairness in pay among crowdworkers. Crowdworkers who are paid more should have higher expectations of how much work they need to perform to receive the payment (Mason and Watts 2009). Expectations set the threshold of what is fair and not fair (Greenberg and Colquitt 2013). Therefore, increases in payment amount should lead to increases in PFP.

Second, perceived fairness in pay should be positively associated with performance quality among crowdworkers. As perceived fairness in pay increases, so should the amount of effort and attention on the part of the crowdworker (Colquitt et al. 2001; Janssen 2001). Increases in the exertion of both physical effort and mental attention have been shown to increase workers' performance in more traditional settings (Jirjahn 2016; Karriker and Williams 2007; Srinivasan, Maruping, and Robert 2010; Srinivasan, Maruping, and Robert 2012). We would also expect increases in physical effort and mental attention to be associated with increases in the performance quality of crowdworkers.

Finally, the payment amount should increase performance quality by increasing PFP among crowdworkers. Ho et al. (2015) identified the importance of ensuring that payment bonuses are big enough to be salient. We employ a similar logic by stating that increases in payment should be big enough to trigger increases in PFP. Increases in payment lead to increases in performance by increasing PFP. Hence the relationship between payment amount and performance quality should occur by increasing PFP.

*H1: PFP mediates the relationship between the payment amount and performance quality in a) the BC task and b) the IMC task.*

PFP should also mediate the relationship between payment amount and a crowdworker's satisfaction. As stated in hypothesis 1, the payment amount should be positively related to PFP. PFP is also positively associated with satisfaction. PFP can influence how much employees enjoy their job (Janssen 2001; Witt and Nye 1992). When employees believe they are paid fairly, they also believe that their work is respected and valued (Colquitt et al. 2013; Karriker and Williams 2007). These beliefs are drivers of job satisfaction (Hopkins and Weathington 2006; Skarlicki and Folger 1997). In fact, the organizational literature has found strong support for the relationship between PFP and job satisfaction (Janssen 2001; McFarlin and Sweeney 1992). This should also be true for crowdworkers.

The influence of payment amount on satisfaction occurs when PFP is increased among crowdworkers. Payment drives satisfaction by increasing workers' belief that they are being treated fairly (Gilliland 1993; Hopkins and Weathington 2006; Skarlicki and Folger 1997). This causes workers to believe that their contribution is respected and valued, leading to more satisfaction ( Robert and You 2013; Skarlicki and Folger 1997; Tekleab, Bartol, and Liu 2005). Therefore, the payment amount should increase satisfaction by increasing perceived fairness in pay among crowdworkers.

*H2: PFP mediates the relationship between the payment amount and satisfaction.*

# Experiment

## Method

We conducted a between-subjects field experiment on Amazon Mechanical Turk (MTurk) with two treatments: high payment and low payment. In the high-payment treatment, participants were told that they would receive a payment of $1.00 for their participation ($7.5/hour based on an 8-minute task), while in the low-payment treatment, participants were informed that they would receive $0.40 (roughly equating to $3/hour). Although it varies, MTurk workers typically view $6.00/hour as a fair payment (Mao et al. 2013). We chose payment conditions that were slightly higher and lower than this amount. To ensure that the participants did not know they received less or more money than other participants for the same task, we deployed only one experimental condition at a time. However, to further ensure worker population similarity across conditions, we deployed all conditions on weekday mornings.

However, to ensure fair treatment and to help create a sustainable crowd work environment, we paid additional compensation after the experiment to reach a compensation rate of $7.5 per hour, which was in accordance with the ongoing payment goals on MTurk ("Fair Payment" 2016). Because we did this after all the data were collected and participants didn't know about this extra compensation at the time they completed the task, we do not believe this influenced the outcome of our study.

## Tasks

To explore the effects of perceived fairness in pay on performance quality and satisfaction, we employed two tasks that were used in previous crowd work studies (e.g., Oppenheimer, Meyvis, and Davidenko 2009; Yin et al. 2013). The first was an effort-responsive button-clicking (BC) task that measured the accuracy of a participant's response. The second was an instructional manipulation check (IMC) task that measured whether the participants paid attention to the instructions. These tasks were chosen for two reasons. First, they represent both an effort- and a non-effort-responsive attention-based task, which allows us to generalize our findings to the emerging importance of task type. Second, these tasks measure performance quality and have been widely used in the literature on crowd work (e.g., Goodman et al. 2013; Horton and Chilton 2010; Yin et al. 2013). Third, they represent two of the most critical task elements found in many other tasks: effort and attention. For example, other commonly used tasks, such as data cleaning and image labeling, involve some degree of both of these elements. Finally, these tasks allow us to better tease apart effort and attention, thereby ruling out any potential confound associated with other tasks that involve some degree of both.

### The button-clicking (BC) task

The BC task was adapted from previous papers (Horton, Rand, and Zeckhauser 2011; Yin et al. 2013). There were three buttons on the screen, one on the top, one in the middle, and one at the bottom. The "target" button was labeled "Click me" and was green, and the other two buttons were labeled "Not me," and one was white and the other orange. Figure 1 shows the interface.

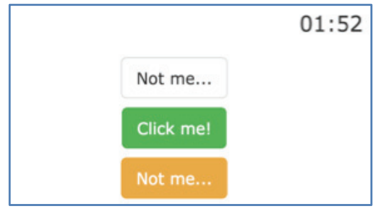


*Figure 1. Interface of the button-clicking task.*

The interface refreshed once a button was clicked and the three buttons randomly changed their position, moving from the top, middle, and bottom. Participants were asked to click the target button as many times as they could in 2 minutes.

### The instructional manipulation check (IMC) task

The IMC task is used to detect satisficing behavior (Oppenheimer et al. 2009). In this task, participants were shown a series of three screens. Each screen had a title on the top followed by a paragraph, a question in the middle with several answer options, and a continue button at the bottom. In the first two screens, they were asked to read a paragraph, which instructed them to make a preference choice among several options. But in the third screen they were instructed to avoid making their preference choice and instead just click the title (see Figure 2). Participants who read and paid attention just clicked the title and those who did not pay attention made their preference choices.

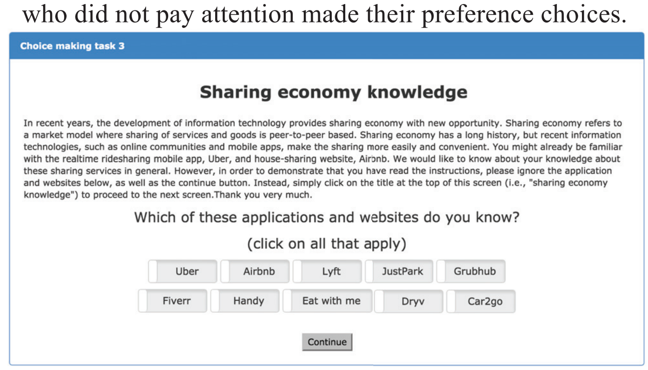


*Figure 2. Interface of the IMC task, subtask 3.*

## Procedure

Participants received consent forms when they arrived. After completing the form, participants were asked to fill out a demographic questionnaire, which included questions about their employment, wages, and trust disposition. Participants then completed the two tasks. To avoid any ordering effects (Newell and Ruths 2016), the order of the BC and IMC tasks was randomized. This resulted in a 2 (low payment vs. high payment) X 2 (task order 1 vs. task order 2) design shown in Table 1. Participants then completed a questionnaire to measure their PFP and satisfaction.

| Treatment | Order 1 | Order 2 |
|---|---|---|
| Low-payment | \$0.40 + IMC first | \$0.40 + BC first |
| High-payment | \$1.00 + IMC first | \$1.00 + BC first |

*Table 1. Treatment conditions*

## Participants

To avoid any effects of culture differences (e.g., Goodman, Cryder, and Cheema 2013), we restricted the experiment to participants located in the United States. A total of 160 participants finished our experiment. After excluding observations with missing data, we had 152 participants in total, with 77 in the low-payment condition and 75 in the high-payment condition. Their age ranged 19–58 years, with a mean of 32.84 years.

## Independent and Dependent Variables

### Perceived fairness in pay (PFP)

To measure PFP, we adapted items using 5-point Likert scale from Colquitt (2001): (1) "My payment reflects the effort I have put into the task," (2) "My payment is appropriate for the work I have completed," and (3) "My payment is justified given my performance." Their Cronbach's alpha was 0.95.

### Clicks on target — performance for the BC task

To measure a participant's performance quality in the button-clicking task, we used the number of clicks on the target button (Yin et al. 2013).

### Title click — performance for the IMC task

We measured performance for the IMC task as either "1" if participants clicked the title on the third screen or "0" if they did not; thus this was a binary outcome variable.

### Total task completion time

We also included total task completion time in our analysis. This was calculated by measuring the time (in seconds) between when the task was accepted and submitted.

### Satisfaction

To measure satisfaction, we adapted three items using 5-point Likert scale from Dennis et al. (1999): (1) "I am satisfied with the task processes," (2) "I feel satisfied with my performance on these tasks," and (3) "Overall, I am satisfied with my experience." The Cronbach's alpha was 0.87.

## Control Variables

We controlled for several factors. We controlled for age, gender, education, average weekly wage from MTurk, and employment status.

Gender was coded as a dummy variable where "0" represented female and "1" represented male. Education was coded as a dummy variable where a college degree or higher was coded as "1" and others "0." Employment was coded as a binary variable, with "1" referring to full-time employment or student status and "0" referring to part-time employment or student status.

We also controlled for disposition to trust, which refers to an individual's general predisposition to trust other people and is akin to a personality trait (Mayer, Davis, and Schoorman 1995). It is an important control variable in the study of PFP (Colquitt, Scott, and LePine 2007). We used items from a 5-point Likert scale to measure disposition to trust (Jarvenpaa, Knoll, and Leidner 1998; Robert, Dennis, and Hung 2009; Schoorman, Mayer, and Davis 1996). Example items included: (1) "Most people can be counted on to do what they say they will do" and (2) "Most people answer personal questions honestly." Cronbach's alpha was 0.90.

# Results

We used SPSS version 22.0 (IBM, Armonk, NY) to test the hypotheses. Table 2 lists the means, standard deviations and correlations of all variables and Table 3 shows the means and standard deviations of the control variables. Our analysis strategy consisted of three sets of statistical tests. First, because we employed a between-subjects field experiment we employed ANCOVA and MANCOVA to test our treatments. Specifically, we used ANCOVA to determine whether the actual payment amount could explain differences in PFP. MANCOVA was used to determine whether pay amount and PFP were associated with differences in outcomes between the two treatments.

Second, because the IMC task had a binary outcome we employed logistic regression. ANCOVA and MANCOVA cannot be used to examine binary outcomes. Third, to test whether increases in actual pay impacted performance quality and satisfaction through PFP (i.e. mediation) we performed a Sobel test. The Sobel test is a statistical analysis to determine whether the effects of one variable on another variable occur through a third, mediating variable (Nguyen, Dabbish, and Kiesler 2015; Sobel 1982). The Sobel test accomplishes this by statistically examining whether the indirect effects associated with the mediation are significant.

Results are shown in Tables 5, 6, and 7. In these tables, Model 1 displays the effects of the control variables. Model 2 shows the effects of the independent variable (payment level), and model 3 includes the mediator (PFP) used to test the mediation effects on each outcome. We also in-

| | Variables | Mean | Std. Dev. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Instruction reading | 0.82 | 0.38 | | | | | | | | | | |
| 2 | Number of clicks on target | 146.55 | 31.93 | 0.01 | | | | | | | | | |
| 3 | Total time | 1130.25 | 794.60 | -0.09 | -0.14 | | | | | | | | |
| 4 | Satisfaction | 4.06 | 0.84 | 0.05 | 0.12 | -0.12 | | | | | | | |
| 5 | Payment level | 0.49 | 0.50 | -0.06 | -0.05 | -0.18* | 0.25** | | | | | | |
| 6 | Gender | 0.49 | 0.50 | -0.09 | 0.07 | -0.07 | -0.23 | -0.03 | | | | | |
| 7 | Age | 32.84 | 8.33 | 0.10 | -0.14 | 0.01 | 0.07 | 0.09 | -0.31*** | | | | |
| 8 | Full-time work&study | 0.55 | 0.50 | -0.15 | -0.10 | -0.03 | -0.14 | -0.13 | 0.08 | -0.12 | | | |
| 9 | Weekly MTurk wage | 5.28 | 1.19 | 0.02 | 0.06 | 0.09 | -0.03 | 0.13 | 0.10 | 0.02 | -0.06 | | |
| 10 | Disposition to trust | 3.19 | 0.70 | 0.16 | -0.08 | -0.03 | 0.33*** | -0.01 | -0.15 | 0.12 | -0.05 | -0.10 | |
| 11 | Perceived fairness in pay | 3.25 | 1.13 | 0.01 | 0.12 | -0.24** | 0.51*** | 0.30*** | -0.14 | -0.06 | 0.06 | -0.09 | 0.23** |

N=152; Significance of correlations: *p< .05; **p< .01; ***p< .001

*Table 2. Means, standard deviations and correlations*

clude total task completion time as an additional analysis because it is often used as important measure of efficiency (e.g., Frøkjær, Hertzum, and Hornbæk 2000).

## Manipulation Check

To check whether our manipulation of high and low payments has the desired effect we conducted a manipulation check. A *t*-test showed that the mean PFP of the high-pay condition (mean = 3.59, standard deviation [SD] = 1.10) was significantly higher than in the low-pay condition (mean = 2.92, SD = 1.08) with t = 3.81, p < 0.001. Therefore, our manipulation was successful.

| Variables | Low-payment condition | | High-payment condition | |
|---|---|---|---|---|
| | Mean | Std. Dev. | Mean | Std. Dev. |
| Gender | 0.51 | 0.50 | 0.48 | 0.50 |
| Age | 32.13 | 8.09 | 33.57 | 8.57 |
| Education | 0.70 | 0.46 | 0.77 | 0.42 |
| Employment | 0.61 | 0.49 | 0.48 | 0.50 |
| Weekly wage (on MTurk) | 5.13 | 1.16 | 5.44 | 1.21 |
| Disposition to trust | 3.25 | 0.86 | 3.22 | 0.89 |

Significance of correlations: *p< .05; **p< .01; ***p< .001
N=152;

*Table 3. Means, standard deviations of control variables*

## Measurement Validity

We conducted a factor analysis. Cross-loadings were under 0.37 and items loaded at or above 0.78 (see Table 4).

| Items | Disposition to trust (DTT) | Perceived fairness in pay (PFP) | Satisfaction |
|---|---|---|---|
| DTT 1 | **0.88** | 0.00 | 0.20 |
| DTT 2 | **0.86** | 0.03 | 0.08 |
| DTT 3 | **0.87** | 0.18 | 0.09 |
| DTT 4 | **0.86** | 0.15 | 0.11 |
| PFP 1 | 0.06 | **0.90** | 0.29 |
| PFP 2 | 0.13 | **0.92** | 0.20 |
| PFP 3 | 0.13 | **0.92** | 0.25 |
| Satisfaction 1 | 0.17 | 0.37 | **0.79** |
| Satisfaction 2 | 0.17 | 0.12 | **0.86** |
| Satisfaction 3 | 0.09 | 0.33 | **0.86** |

Extraction method: principal component analysis.
Rotation method: Varimax with Kaiser Normalization.

*Table 4. Factor loadings*

## Mediation, Indirect Effects and the Sobel Test

Originally, mediation was defined by Baron and Kenny's three steps for mediation (Baron and Kenny 1986). The three steps are: step 1, the independent variable should be a predictor of the dependent variable; step 2, the independent variable should be a predictor of the mediator variable; step 3, the mediator variable should be a predictor of the dependent variable, in the presence of the independent variable. Full mediation is said to occur if the independent variable is no longer a significant predictor of the dependent variable in the presence of the mediator variable. Partial mediation is said to occur when the independent variable is still a significant predictor of the dependent variable in the presence of the mediator variable.

However, scholars have reassessed the requirements for mediation in response to several criticisms. First, scholars have now dropped Baron and Kenny's (1986) step 1 requirement. Scholars have found that the independent varia-

| Variables | Number of clicks on target | | | Satisfaction | | | Total time (in seconds) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 1 | Model 2 | Model 3 | Model 1 | Model 2 | Model 3 |
| **Control variables** | | | | | | | | | |
| Gender | 1.54 | 1.51 | 3.27 | -0.31* | -0.31* | -0.21 | -139.22 | -141.27 | -187.48 |
| Age | -0.52 | -0.51 | -0.40 | 0.00 | -0.01 | 0.00 | -0.75 | 0.41 | -2.51 |
| Education | -4.89 | -4.55 | -4.66 | 0.04 | 0.00 | 0.00 | -256.18 | -227.94 | -224.92 |
| Employment | -6.98 | -7.43 | -8.68 | -0.20 | -0.15 | -0.22 | -8.35 | -45.15 | -12.15 |
| Weekly wage (on MTurk) | 1.44 | 1.62 | 2.08 | 0.01 | -0.01 | 0.01 | 66.85 | 81.59 | 69.52 |
| Disposition to trust | -2.08 | -2.11 | -3.93 | 0.29*** | 0.29*** | 0.19** | -30.21 | -32.44 | 15.34 |
| **Independent variable** | | | | | | | | | |
| Payment level | | -3.60 | -7.82 | | 0.40** | 0.17 | | -298.19* | -187.52 |
| **Mediator** | | | | | | | | | |
| Perceived fairness in pay | | | 5.61* | | | 0.31*** | | | -147.19* |

N=152; Significance of coefficients: *p< .05; **p< .01; ***p< .001

*Table 5. Results of MANCOVA analyses*

ble can still indirectly impact the dependent variable through the mediator variable, even when it does not directly impact the dependent variable (Shrout and Bolger 2002). Second, scholars have now developed formal statistical tests of mediation. The Sobel test is widely used as a formal statistical mediation test (Hayes 2009).

| Variables | Title Click | | |
|---|---|---|---|
| | Model 1 (Odds ratio) | Model 2 (Odds ratio) | Model 3 (Odds ratio) |
| **Control variables** | | | |
| Gender | 1.36 | 1.39 | 1.38 |
| Age | 1.02 | 1.02 | 1.02 |
| Education | 2.29 | 2.17 | 2.17 |
| Employment | 2.00 | 2.14 | 2.16 |
| Weekly wage (on MTurk) | 1.07 | 1.10 | 1.10 |
| Disposition to trust | 1.46 | 1.46 | 1.45 |
| **Independent variable** | | | |
| Payment level | | 1.54 | 1.58 |
| **Mediator** | | | |
| Perceived fairness in pay | | | 1.04 |

N=152; Significance of coefficients: *p< .05; **p< .01; ***p< .001

*Table 6. Results of the logistic regression predicting title click*

## Mediation: Empirical Test of Indirect Effects

To test our hypotheses, we performed three steps. First, we tested the direct effects of actual payment and PFP on each outcome variable. Second, we determined whether actual payment increased PFP. Finally, we conducted a Sobel test to determine whether the indirect effects of payment amount through PPF were significant.

Step 1, the payment amount did not impact performance quality in the BC task (B = –3.60, standard error [SE] = 5.30, t = –0.68, p = 0.50) as seen in Table 5 or performance quality in the IMC task (B = 0.43, SE = 0.46, 95% CI [0.63, 3.78], odds ratio = 1.54, p = 0.34) as seen in Table 6. This supports previous research on the impact of payment amount in the crowdsourcing literature (Mason and Watts 2009; Yin et al. 2013). However, actual payment did predict satisfaction (B = 0.40, SE = 0.13, t = 3.19, p < 0.01) as seen in Table 5.

| Variables | Perceived fairness in pay | |
|---|---|---|
| | Model 1 | Model 2 |
| **Control variables** | | |
| Gender | -0.32 | -0.31 |
| Age | -0.02 | -0.02 |
| Education | 0.09 | 0.02 |
| Employment | 0.13 | 0.22 |
| Weekly wage (on MTurk) | -0.05 | -0.08 |
| Disposition to trust | 0.32** | 0.33** |
| **Independent variable** | | |
| Payment level | | 0.75*** |

N=152; Significance of coefficients: *p< .05; **p< .01; ***p< .001

*Table 7. Results of ANCOVA analyses*

Step 2, we examined whether PFP was associated with increases in performance quality and satisfaction. PFP was associated with increases in performance quality for the BC task (B = 5.61, SE = 2.53, t = 2.22, p < 0.05) as seen in

Table 5, but not for the IMC task (B = 0.04, SE = 0.21, 95% CI [0.68, 1.57], odds ratio = 1.04, $p = 0.86$) as seen in Table 6. However, PFP did lead to increases in satisfaction (B = 0.31, SE = 0.06, t = 5.60, $p < 0.001$) see Table 5.

Next, we examined whether payment amount impacted PFP. To accomplish this, we conducted an ANCOVA. Results are shown in Table 7. Payment amount was significantly related to PFP (B = 0.75, SE = 0.17, t = 4.37, $p < 0.001$). Next we discuss the test results of the mediation effects related to each hypothesis.

#### Hypothesis 1a: Performance quality in the BC task

The Sobel test was conducted to evaluate hypothesis 1a. Results of the Sobel test indicate that the indirect effect of actual payment on the performance quality of the button-clicking task was mediated through PFP (Z = 1.98, $p < 0.05$). H1a was supported.

#### Hypothesis 1b: Performance quality in the IMC task

Hypothesis 1b asserts that PFP mediates the relationship between the actual payment and the performance quality in the IMC task. Sobel test results showed that H1b was not supported with Z = 0.19, $p = 0.85$.

#### Hypothesis 2: Satisfaction

The Sobel test indicated that the indirect effect of payment amount on satisfaction was mediated through PFP (Z = 3.45, $p < 0.001$), supporting H2.

#### Additional analysis: Total task completion time

Additionally, we tested whether PFP mediated the relationship between actual pay and the total task time. As reported in Table 5, increases in PFP (B = –147.19, SE = 62.10, t = –2.37, $p < 0.05$) were associated with decreases in total time. The Sobel test was significant (Z = 2.08, $p < 0.05$). Increases in actual payment led to faster total task completion time by increasing PFP.

### Summary of Results

Results of this study can be organized into the two major findings. First, PFP mediated the impact of payment amount on performance quality in the button-clicking task but not in in the IMC task. In fact, neither payment amount nor PFP was significantly associated with performance quality in the IMC task. Like prior researchers, we found that the benefits associated with more pay vary by task type. Higher pay increased performance quality in effort-responsive tasks but not in non-effort-responsive attention-based tasks.

There are at least two potential explanations for why PFP was positively associated with performance quality in the BC task and not the IMC task. One explanation is the task completion time. Crowdworkers who were paid more might have been too motivated to complete the task too quickly. Increases in motivation might help in the BC task but not in the IMC task. For example, crowdworkers who were paid more and had higher levels of perceived fairness in pay performed the task much faster (mean = 978.44, SD = 652.11) than their counterparts (mean = 1,252.33, SD = 883.83). We examined the correlations between task time and performance quality for the BC task (correlation coefficient = –0.14, p = .08) and IMC task (correlation coefficient = –0.09, p = .26). Both were not significant at the .05 level. Another explanation could be a lack of power. We conducted a power analysis. Power analyses for H1a = .99, and H1b = .83 both indicated high statistical power.

Second, PFP mediated the impact of payment amount on satisfaction (supporting H2) and total task completion time (an additional analysis). Previous studies investigating the impact of payment magnitude on task time among crowdworkers have found no significant effect (Rogstadius et al. 2011; Yin et al. 2013). We could not find any study examining the impact of payment amount on satisfaction among crowdworkers. This suggests that PFP could be used to help understand when pay increases can lead to other types of crowdworker outcomes.

## Discussion

Our goal was to determine whether PFP can help us better understand the relationship between pay and crowdworkers' performance quality. To that end, our results demonstrate the potential usefulness of the inclusion of PFP in the study of pay and performance quality in crowdsourcing.

### Contributions and Implications

First, in this paper we propose an alternative view of crowdworkers, their motivations, and the use of financial incentives to increase their performance. In doing so, we acknowledge that crowdworkers are not human CPUs that simply react to increases and decreases in the inputs given to them. Instead, they are individuals with a sense of fairness who can react irrationally and emotionally based on their perceptions of their treatment.

Second, in doing so this study broadens the discussion on pay and crowd work to include satisfaction and task time. When crowdworkers thought they were paid fairly, not only did they work faster, but they also were more satisfied. There has been much discussion on the importance of fair pay and the need for better treatment of crowdworkers (e.g., Irani et al. 2013; Martin et al. 2014), along with the discussion on creating a better workplace for crowdworkers (Kittur et al. 2013). Understanding the role of PFP in promoting satisfaction is an important step forward.

Third, we identified an important mediator that links the payment amount and performance quality in effort-responsive tasks. Our results showed that pay increases had an indirect effect on performance quality by increasing an individuals' PFP. To improve performance quality, future research can further explore other predictors of PFP. For example, the relationship quality between employees and their employer is an important predictor of PFP (Colquitt et al. 2013). It is not clear how such findings can be directly translated to a crowdsourcing context. On one hand, Turkers may view requestors as their employer. Turkers are aware of a requester's reputation, either through direct experience or through reputation systems. Turkers may decide to accept a task based on this information. From this point of view, the requester is the employer. On the other hand, Turkers normally perform work for many different requestors, and in at least some cases their relationship with a requestor could be a one-time occurrence.

However, MTurk sets up the overall working conditions between the worker and the requestor. Workers maintain their relationship with MTurk no matter which requester they work for. Their relationship with MTurk is also much more permanent than their relationship with a particular requestor. Can this lead Turkers to view MTurk as their employer? Can Turkers also view both as their employer? We simply do not know. To the best of our knowledge no research has been conducted on this particular topic in the context of crowdsourcing. Future research is needed in this area and others to begin to understand what other factors are likely to drive PFP in the context of crowdsourcing.

Fourth, this paper offers a possible explanation for the seemingly conflicting results regarding the relationship between payment amount and performance quality. Some researchers found that increases in pay did not lead to increases in crowdworker performance quality (e.g., Buhrmester et al. 2011; Mason and Watts 2009), while others found payment increases did have direct effects on performance quality (Ho et al. 2015).

We found that the payment amount influenced performance quality via the mediator of PFP. Ho et al. (2015) suggested that increases in pay may only lead to increases in performance quality when they are big enough to be salient. Our results help to explain this argument. Payment increases should have direct effects on performance quality when they are big enough to be salient. However, when pay increases are not large enough to be salient, increases in pay will not have a direct effect on performance quality. Individuals with different backgrounds and personalities may differ on what amount is big enough to be salient. Therefore, the same pay increase does not necessarily translate to being big enough for every individual. The inclusion of PFP may help us capture these differences and provides the link between pay increases and increases in performance quality for effort-responsive tasks.

Fifth, PFP did not help us understand why increases in pay do not lead to increases in performance quality in non-effort-responsive tasks or in our case *attention-based* tasks. Our results are similar to those in other research: we found that increases in pay lead to increases in performance in effort-responsive tasks only (see Ho et al. 2015). It would seem that increases in pay lead to more effort but not more attention. One explanation is that attention tasks are much more dependent on training. Crowdworkers might have to be trained to become better at attention tasks. Future studies could examine whether training is more or less important for attention tasks relative to effort tasks.

Finally, this study has implications for the design of crowdsourcing platforms. Crowdsourcing systems can be used to facilitate PFP by convincing crowdworkers that the payment is fair relative to the task. There are two ways to accomplish this objective. One approach is to explain why the payment is fair relative to the task requirements. Crowdworkers might be more likely to believe payment is fair when they understand how the amount was determined. This could be accomplished by having the system display the breakdown of the overall task into subtasks. The payment relative to the amount of work anticipated for each subtask could be displayed alongside each subtask. This would provide greater transparency on how the payment amount was derived. This, in turn, could be used to facilitate perceived fairness in pay without actually increasing payment.

Another way would be to add a feature to the system that calculates the active time participants spend to complete the task. Amazon Mechanical Turk currently provides workers with the average completion time for a task. But this number is not accurate because workers rarely work straight through a task and often get distracted by other activities (e.g., completing other tasks). Therefore, the listed average task time is likely to be an overestimation of the actual time required to complete the task. This overestimation is likely to lead crowdworkers to believe that payment should be higher relative to the actual time needed to complete the task. We suggest developing a tool to calculate the active time that a worker actually spends on a task. All other things being equal, a more accurate estimate of the actual task time is likely to increase perceptions of fairness in pay.

## Limitations

This study is limited in several ways. First, we only conducted the experiment in one crowdsourcing platform. Second, we restricted our participants to the United States in order to rule out potential cultural differences (Shaw, Horton, and Chen 2011). One of our next steps is to test whether our results hold in other populations. Third, in order to make the study comparable to previous studies on performance (e.g., Mao et al. 2013; Yin et al. 2013), we didn't set restrictions on approval rate. Future work needs to be done on Master Turkers. Fourth, the two tasks were chosen to better tease apart the effect on attention-based and effort-responsive tasks. However, future study could deploy tasks that are more realistic and commonly used in crowdsourcing platforms. Finally, this study didn't include intrinsic motivations. It would be interesting to see the interaction effects of the magnitude of financial incentives, perceived fairness in pay, and intrinsic motivations on task performance and satisfaction.

## Conclusion

Via the lens of perceived fairness in pay, this paper sought to extend our understanding of the relationship between financial incentives and crowdworkers' performance. This study highlights the importance of perceptions of fairness in pay and indicates potential design applications to improve fairness perceptions and task performance in crowdsourcing markets.

## Acknowledgments

We thank the reviewers for their insightful feedback and the participants for their participation. We especially thank Chanda Phelan, Carol Moser, Andrea Barbarin and Fangzhou Zhang for providing helpful comments on previous versions of this document. We also gratefully acknowledge the financial support from the National Science Foundation grant CHS-1617820.

## References

Adams, J. S. 1965. "Inequity In Social Exchange." In Advances in Experimental Social Psychology, edited by Leonard Berkowitz, 2:267–99. Academic Press.

Baron, R. M. and Kenny, D. A. 1986. "The Moderator–mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." Journal of Personality and Social Psychology 51 (6): 1173.

Buhrmester, M.; Kwang, T.; and Gosling, S. D. 2011. "Amazon's Mechanical Turk a New Source of Inexpensive, yet High-Quality, Data?" Perspectives on Psychological Science 6 (1): 3–5.

Camerer C.; Babcock L.; Loewenstein G.; and Thaler R. 1997. "Labor Supply of New York City Cabdrivers: One Day at a Time." The Quarterly Journal of Economics 112 (2): 407–441.

Colquitt, J. A. 2001. "On the Dimensionality of Organizational Justice: A Construct Validation of a Measure." Journal of Applied Psychology 86 (3): 386.

Colquitt, J. A.; Scott, B. A.; and LePine, J. A. 2007. "Trust, Trustworthiness, and Trust Propensity: A Meta-Analytic Test of Their Unique Relationships with Risk Taking and Job Performance." Journal of Applied Psychology 92 (4): 909.

Colquitt, J. A.; Scott, B. A.; Rodell, J. B.; Long, D. M.; Zapata, C. P.; Conlon, D. E.; and Wesson, M. J. 2013. "Justice at the Millennium, a Decade Later: A Meta-Analytic Test of Social Exchange and Affect-Based Perspectives." Journal of Applied Psychology 98 (2): 199.

Dennis, A. R.; Aronson, J. E.; Heninger, W. G.; and Walker, E. D. 1999. "Structuring Time and Task in Electronic Brainstorming." MIS Quarterly, 95–108.

"Fair Payment." 2016. http://wiki.wearedynamo.org/index.php?title=Fair_payment#What_is_ethical_pay_for_Turkers_in_studies.3F.

Folger, R. 1977. "Distributive and Procedural Justice: Combined Impact of Voice and Improvement on Experienced Inequity." Journal of Personality and Social Psychology 35 (2): 108.

Folger, R.; and Konovsky, M. A. 1989. "Effects of Procedural and Distributive Justice on Reactions to Pay Raise Decisions." Academy of Management Journal 32 (1): 115–30.

Frøkjær, E.; Hertzum, M. and Hornbæk, K. 2000. "Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated?" In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 345–352. ACM.

Gilliland, S. W. 1993. "The Perceived Fairness of Selection Systems: An Organizational Justice Perspective." Academy of Management Review 18 (4): 694–734.

Goodman, J. K.; Cryder, C. E.; and Cheema, A. 2013. "Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples." Journal of Behavioral Decision Making 26 (3): 213–224.

Greenberg, J.; and Colquitt, J. A. eds. 2013. Handbook of Organizational Justice. Psychology Press.

Hayes, A. F. 2009. "Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium." Communication Monographs 76 (4): 408–420.

Ho, C.J., Slivkins, A., Suri, S. and Vaughan, J.W., 2015, May. Incentivizing high quality crowdwork. In Proceedings of the 24th International Conference on World Wide Web, 419-429. ACM.

Hopkins, S. M.; and Weathington, B. L. 2006. "The Relationships Between Justice Perceptions, Trust, and Employee Attitudes in a Downsized Organization." The Journal of Psychology 140 (5): 477–98.

Horton, J. J.; and Chilton, L. B. 2010. "The Labor Economics of Paid Crowdsourcing." In Proceedings of the 11th ACM Conference on Electronic Commerce, 209–218. EC '10. New York, NY, USA: ACM.

Horton, J. J.; Rand, D. G.; and Zeckhauser, R. J. 2011. "The Online Laboratory: Conducting Experiments in a Real Labor Market." Experimental Economics 14 (3): 399–425.

Irani, L. C.; and Silberman, M. Six. 2013. "Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk." In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 611–620. CHI '13. New York, NY, USA: ACM.

Janssen, O. 2001. "Fairness Perceptions as a Moderator in the Curvilinear Relationships Between Job Demands, and

Job Performance and Job Satisfaction.” Academy of Management Journal 44 (5): 1039–50.

Jarvenpaa, S. L.; Knoll, K.; and Leidner, D. E. 1998. “Is Anybody out There? Antecedents of Trust in Global Virtual Teams.” Journal of Management Information Systems 14 (4): 29–64.

Jirjahn, U. 2016. “Performance Pay and Productivity: A Note on the Moderating Role of a High-Wage Policy.” Managerial and Decision Economics.

Kahneman, D. 2003. “Maps of Bounded Rationality: Psychology for Behavioral Economics.” The American Economic Review 93 (5): 1449–1475.

Karriker, J. H.; and Williams, M. L. 2007. “Organizational Justice and Organizational Citizenship Behavior: A Mediated Multifoci Model.” Journal of Management.

Kittur, A.; Jeffrey V. N.; Bernstein, M.; Gerber, E.; Shaw, A.; Zimmerman, J.; Lease, M.; and Horton, J. 2013. “The Future of Crowd Work.” In Proceedings of the 2013 Conference on Computer Supported Cooperative Work, 1301–1318. CSCW ’13. New York, NY, USA: ACM.

Mao, A.; Kamar, E.; Chen, Y.; Horvitz, E.; Schwamb, M. E.; Lintott, C. J.; and Smith A. M. 2013. “Volunteering versus Work for Pay: Incentives and Tradeoffs in Crowdsourcing.” In First AAAI Conference on Human Computation and Crowdsourcing.

Martin, D.; Hanrahan, B. V.; O’Neill, J.; and Gupta N. 2014. “Being a Turker.” In Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, 224–235. CSCW ’14. New York, NY, USA: ACM.

Mason, W.; and Watts, D. J. 2009. “Financial Incentives and the ‘Performance of Crowds.’” In Proceedings of the ACM SIGKDD Workshop on Human Computation, 77–85. HCOMP ’09. New York, NY, USA: ACM.

Mayer, R. C.; Davis, J. H.; and Schoorman, F. D. 1995. “An Integrative Model of Organizational Trust.” Academy of Management Review 20 (3): 709–734.

McFarlin, D. B.; and Sweeney, P. D. 1992. “Research Notes. Distributive and Procedural Justice as Predictors of Satisfaction with Personal and Organizational Outcomes.” Academy of Management Journal 35 (3): 626–37.

Newell, E.; and Ruths, D. 2016. “How One Microtask Affects Another.” In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 3155–3166. CHI ’16. New York, NY, USA: ACM.

Nguyen, D. T.; Dabbish, L. A.; and Kiesler, S. 2015. “The Perverse Effects of Social Transparency on Online Advice Taking.” In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, 207–217. CSCW ’15. New York, NY, USA: ACM.

Oppenheimer, D. M.; Meyvis, T.; and Davidenko, N. 2009. “Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power.” Journal of Experimental Social Psychology 45 (4): 867–872.

Robert, L. P.; Denis, A. R.; and Hung, Y. C. 2009. “Individual Swift Trust and Knowledge-Based Trust in Face-to-Face and Virtual Team Members.” Journal of Management Information Systems 26 (2): 241–279.

Robert, L., and You S. 2013. “Are You Satisfied yet? Shared Leadership, Trust and Individual Satisfaction in Virtual Teams.” In iConference 2013 Proceedings, 461–466.

Rogstadius, J.; Kostakos, V.; Kittur, A.; Smus, B.; Laredo, J.; and Vukovic, M. 2011. “An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets.” ICWSM 11: 17–21.

Saunders, M. N. K.; and Thornhill, A. 2003. “Organisational Justice, Trust and the Management of Change: An Exploration.” Personnel Review 32 (3): 360–75.

Schoorman, F. D.; Mayer, R. C.; and Davis, J. H. 1996. “Empowerment in Veterinary Clinics: The Role of Trust in Delegation.” In 11th Annual Meeting of the Society for Industrial and Organizational Psychology, San Diego, CA.

Shaw, A. D.; Horton, J. J.; and Chen, D. L. 2011. “Designing Incentives for Inexpert Human Raters.” In Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work, 275–284. CSCW ’11. New York, NY, USA: ACM.

Shrout, P. E.; and Bolger, N. 2002. “Mediation in Experimental and Nonexperimental Studies: New Procedures and Recommendations.” Psychological Methods 7 (4): 422.

Skarlicki, D. P.; and Folger, R. 1997. “Retaliation in the Workplace: The Roles of Distributive, Procedural, and Interactional Justice.” Journal of Applied Psychology 82 (3): 434–43.

Sobel, M. E. 1982. “Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models.” Sociological Methodology 13 (1982): 290–312.

Srinivasan, S.; Maruping L. M.; and Robert L. P. 2010. “Mechanisms Underlying Social Loafing in Technology Teams: An Empirical Analysis.” In AIS.

Srinivasan, S.; Maruping L. M.; and Robert L. P. 2012. “Idea Generation in Technology-Supported Teams: A Multilevel Motivational Perspective.” In System Science (HICSS), 2012 45th Hawaii International Conference on, 247–256. IEEE.

Tekleab, A. G.; Bartol, K. M.; and Liu, W. 2005. “Is It Pay Levels or Pay Raises That Matter to Fairness and Turnover?” Journal of Organizational Behavior 26 (8): 899–921.

Tversky, A.; and Kahneman D. 1992. “Advances in Prospect Theory: Cumulative Representation of Uncertainty.” Journal of Risk and Uncertainty 5 (4): 297–323.

Witt, L. A.; and Nye, L. G. 1992. “Gender and the Relationship between Perceived Fairness of Pay or Promotion and Job Satisfaction.” Journal of Applied Psychology 77 (6): 910.

Yin, M.; and Chen, Y. 2016. “Predicting Crowd Work Quality under Monetary Interventions.”

Yin, M.; Chen, Y.; and Sun, Y. 2013. “The Effects of Performance-Contingent Financial Incentives in Online Labor Markets.”