11-29-2018

Antecedents of Perceived Fairness in Pay for Microtask Crowdwork

Paul Alpar
University at Marburg, alpar@wiwi.uni-marburg.de

Lars Osterbrink
University at Marburg, lars.osterbrink@wiwi.uni-marburg.de

ANTECEDENTS OF PERCEIVED FAIRNESS IN PAY FOR MICROTASK CROWDWORK

Research in Progress

Alpar, Paul, University of Marburg, Marburg, Germany, alpar@uni-marburg.de

Osterbrink, Lars, University of Marburg, Marburg, Germany, osterbrl@uni-marburg.de

Abstract

Crowdwork has become a powerful tool for businesses and researchers to get work done in a fast, convenient, and cheap way. On one hand, literature suggests that high data quality can even be achieved with poor payment which has become common practice in crowdwork. On the other hand, recent research and ethical considerations suggest that poor payments and especially a low perceived fairness in pay may come at a price. Crowdworkers may put less effort in a task, stop working for a business/researcher, or even leave the crowdsourcing platform entirely. Therefore, we develop a model in this paper to understand how perceived fairness in pay is formed before task execution. If it is only measured after task execution, we miss the "voice" of those who did not even attempt to take on a task. Therefore, we test the effect of perceived fairness in pay on actual task execution and whether it changes after task execution.

Keywords: Crowdwork, Perceived Fairness in Pay, Mechanical Turk, MTurk.

1 Introduction

Crowdsourcing has become a popular social and economic activity. Some of the commercial uses are the collection of money for start-up activities (crowdfunding, Belleflamme et al., 2014), the inclusion of firm outsiders into product development (open innovation, Afuah and Tucci, 2012), or the accomplishment of small jobs by independent workers (microtask crowdsourcing, Deng et al., 2016). This paper concentrates on the latter form of crowdsourcing. Microtask-crowdsourcing platforms like Crowdflower.com, Clickworker.com and the leading platform MTurk.com (Amazon Mechanical Turk, MTurk) are meanwhile utilized by both businesses and researchers. Businesses use MTurk as a way to outsource various tasks to a scalable and always available online workforce (e.g., Ford et al., 2015; Schmidt and Jettinghoff, 2016). For researchers, MTurk primarily represents a fast, convenient and cheap pool of research participants (e.g., Buhrmester et al., 2011; Berinsky et al., 2012; Steelman et al., 2014). Originally, MTurk was not created for academic research but for "human computation" tasks that were easy for humans but difficult for computers to solve like tagging pictures or transcribing audio (Mason and Suri, 2012). However, researchers experimented with MTurk, tested the generated data many times regarding representativeness, reliability, and validity and found MTurk samples to be a viable alternative to student samples or online panels (Lowry et al., 2016; Sheehan, 2017). In general, payments on MTurk are rather low for western industrialized countries like the US (Deng et al., 2016), but there are great differences between employers (Silberman and Irani, 2016). An employer can reject any work for any reason or no reason at all without interference by Amazon (Sheehan, 2017).

On MTurk, every employer has to set a compensation for completing a given task. Additionally, an employer can deliberately give bonus payments for, e.g., extraordinary good work. Setting the "right" amount of money for a task is not a trivial decision. Researchers, for example, are urged to pay a "fair" wage for participation in their surveys and experiments (e.g., Deng et al., 2016; Goodman and Paolacci, 2017). Some researchers refer to the *Guidelines for Academic Requesters* (2017) which were created and are occasionally updated by researchers and MTurk workers. They state that payments below federal minimum wage in the US ($7.25/h; ~$0.12/min in 2018) constitute underpayment for US crowdworkers. 10c/min is considered as borderline acceptable and it is often simply used as a “fair” target pay rate by many researchers (e.g., Daly and Nataraajan, 2015; Chandler and Shapiro, 2016; Casey et al., 2017). Nevertheless, some researchers pay remarkably below the US minimum wage (Fort et al., 2011). High pay increases data collection speed (Mason and Watts, 2009; Buhrmester et al., 2011; Berinsky et al., 2012; Benson et al., 2015), but according to many studies not the data quality (Mason and Watts, 2009; Buhrmester et al., 2011; Goodman et al., 2013; Litman et al., 2015). For many studies, it is not crucial how long the data collection takes, e.g., 24 or 48 hours. Therefore, one could consider the payment issue primarily as a question of ethics (Chandler and Shapiro, 2016). A recent study by Ye et al. (2017) supports this view for attention-based tasks but not for effort-responsive tasks. In the latter case, high pay increases the crowdworkers’ perceived fairness in pay (PFP) which increases performance quality and overall crowdworker satisfaction. Additionally, salient performance-based payments (via bonus payments on MTurk) can increase performance for effort-responsive tasks (Ho et al., 2015).

While PFP may have important consequences, its antecedents have not been studied yet in the context of microtask crowdsourcing. Researchers have studied fairness perceptions in many organizational contexts, for example, employment decisions (Gilliland, 1993), salary raises (Folger and Konovsky, 1989), and remuneration of salespeople (Ramaswami and Singh, 2003). Recently, closer to our context, issues of payment fairness have been studied in crowdsourcing for innovation (Franke et al., 2013). Although platforms for crowdsourcing of microtasks partly operate in a similar manner, there are also a number of significant differences which require specific research as explained later. Another issue that requires careful analysis is the temporal interplay of PFP and its consequences. PFP is probably affected by several work and crowdworker related characteristics in addition to the pay rate itself. We infer these characteristics from the above cited literature and specifics of microtask crowdwork. To our knowledge, there is no study yet that investigates the antecedents of PFP for microtask crowdworkers. Hence, our study addresses the following research questions (RQ):

- RQ 1. How do crowdworkers form their perception of fairness in pay?
- RQ 2. How strong does their perception of fairness in pay influence their decision to perform an offered task?
- RQ 3. Does their perception of fairness in pay change after task execution?

Answers to these questions should be of interest to theory and practise of pay setting in microtask crowdsourcing. In this paper, we do not analyze a crowdsourcing platform as an IT artefact but like many other papers in IS research, e.g., those about adoption of IS, we analyze a specific aspect of its use by crowdworkers. In our model (see below), we use variables that are specific to the platform, its use and the work offered there.

In the next section, we discuss microtask crowdsourcing in more detail and some research findings related to it. In the same section, we also look at the specific subject of our paper, PFP, and current findings on it. Our hypotheses and research model are described in section 3. This is followed by a description of the experiments we intend to conduct to collect the data for hypotheses tests in section 4. Finally, we summarize the current status of our research.

2 Theoretical Background

2.1 Microtask Crowdsourcing

The term crowdsourcing was first introduced by Howe (2006) and describes a new way of accessing a large undefined crowd for, e.g., projects or research on a voluntary basis. There are four essential elements in crowdsourcing: the crowd, the crowdsourcer, the crowdsourced task, and the crowdsourcing platform (Hosseini et al., 2015). Microtask crowdsourcing is arranged via a crowdsourcing platform where crowdsourcer (job requester) post microtasks (crowdwork) and set a monetary reward for completing these microtasks. Then, individuals from the crowd (crowdworker) can voluntarily complete these microtasks and receive the monetary reward (Deng et al., 2016). Although there are several crowdsourcing platforms (e.g., Crowdflower.com, Clickworker.com), MTurk by Amazon is by far the most popular platform for research regarding microtask crowdsourcing. MTurk refers to a microtask as a “human intelligence task” (HIT). In most cases, special training and/or knowledge is not necessary for completing microtasks which can be tagging pictures, transcriptions, surveys or experiments (Brawley and Pury, 2016). There may be cases where crowdworkers maintain a long relationship with a job requester (e.g., in audio transcription), but all microtasks are based on short-term arrangements and neither the job requester nor the crowdworker can force participation in future microtasks (Brawley and Pury, 2016). In contrast to open innovation where only the winner or the best participants receive a relatively large monetary reward and uncertainty of return is high (Franke et al., 2013; Boudreau et al., 2011), almost all crowdworkers on MTurk get paid a small amount of money after completing work. Note, that job requesters on MTurk do not have to specify the estimated HIT completion time but they often do it in the task description. In addition, visibility of individual contribution in open innovation is high and often directly attributable to return to the firm organizing the contest (Franke et al., 2013). In microtask crowdwork, the contribution is often invisible (e.g., searching, clicking, rating, categorizing) and the return to the firm remains invisible for outsiders, even if the individual contribution is identifiable like in the case of a product review.

The demographics and social background of MTurk crowdworkers have been studied many times (Chandler and Shapiro, 2016). With nearly 10,000 respondents, a recent study by Casey et al. (2017) found out that the average US MTurk crowdworker is much younger (33.51 years old) than the general population (47.01) and the male/female distribution is almost the same (~50%). Furthermore, white/caucasian are overrepresented (82.9% on MTurk vs. 73.6% in general population) and Latino heavily underrepresented (5.5% vs. 17.1%). Regarding socioeconomic characteristics, Casey et al. (2017) show that US MTurk crowdworkers are equally full-time employed (~50%), more likely to work part-time (15.7% vs. 10.8%), more likely students (11.9% vs. 6.4%) and less likely to be retired (2.2% vs. 15.4%), but their unemployment rate is much higher (9.4% vs. 4.8%). Surprisingly, almost 57,6% of the US MTurk crowdworkers have at least a 2-year college degree compared to 38,2% in the general population (Casey et al., 2017). These findings are consistent with previous studies (e.g., Berinsky et al., 2012; Paolacci et al., 2010).

Although crowdworkers can be intrinsically motivated to participate in crowdwork (Ipeirotis, 2010; Buhrmester et al., 2011; Mason and Watts, 2009), most crowdworkers’ primary motivation seems to be financial compensation (e.g., Paolacci et al., 2010; Silberman et al., 2010; Behrend et al., 2011; Kaufmann et al., 2011; Martin et al., 2013; Ye et al., 2017). Crowdworkers choose tasks according to their personal preferences (Ipeirotis, 2010). Not only the payment but also the reputation of the job requester can be an important factor for choosing a task (Brawley and Pury, 2016). Crowdworkers prefer job requesters with a good reputation over job requesters with a bad reputation (Schmidt and Jettinghoff, 2016). The reputation can be inferred from personal past experiences with the job requester or from forums for crowdworkers like turkernation.com, mturkforum.com, or Turkopticon.ucsd.edu. Turkopticon allows crowdworkers to review job requesters similar to the product reviews on Amazon. Presumably, more than half of all crowdworkers (Benson et al., 2015) and almost all professional crowdworkers on MTurk utilize Turkopticon to select job requesters to work for (Silberman and Irani, 2016). Crowdworkers can ignore or stop working for any job requester and they are free to leave MTurk at any time (Durward et al., 2016).

2.2 Perceived Fairness in Pay

2.2.1 The Concept

The issue of fairness in pay in microtask crowdsourcing is receiving increased attention (e.g., Schulze et al., 2012; Deng et al., 2016; Ye and Kankanhalli, 2017; Ye et al., 2017). We follow the description of PFP as a person's belief that she is fairly compensated for her work (Ye et al., 2017). Most organizational justice models only consider PFP ex post (Gilliland, 1993; Cohen-Charash and Spector, 2001). Measurement of PFP after task execution (Ye et al., 2017) considers real effort in PFP measurement. However, this way only those participants can be observed who decided to perform the task. People who had the feeling that the pay would not fairly compensate for their work do not even start the task and are missing in the above analysis. When measuring PFP after work completion to calculate its influence on work quality or other outcome variables, another problem exists: it is not clear when the feeling of unfair payment arises. If this happens only after work completion, then the work is already done and cannot affect the quality of work anymore. If it happens before, it may start to affect the quality of work but we do not know when it happened. Therefore, PFP should be assessed ex ante if all people should be included who get the chance to participate (Franke et al., 2013). Schulze et al. (2012) suggest a construct called "payment fit" as an indirect antecedent of "work intention" in microtask crowdworking. The construct itself does not have antecedents and its psychometric quality is not known since its test has not been reported yet, which is also true for the complete research model (Schulze et al., 2012). What drives payment fit is not researched.

In the case of microtask crowdsourcing on MTurk, there are cases where PFP is very low but irrelevant because HITs are paid almost nothing and get done nevertheless. Buhrmester et al. (2011) demonstrate this with a very short HIT asking just for gender and age. Within 33 hours 500 responses were collected for 1 cent each. They conclude that crowdworkers are not primarily driven by financial incentives but by some kind of intrinsic motivation. However, there might be another simple reason why crowdworkers complete small HITs for almost no compensation. Many good-paying HITs from good job requesters (organizations or researchers) require an approval rate of 95% or above (Casler et al., 2013; Brawley and Pury, 2016; Sheehan, 2017). The approval rate is the ratio of approved to all submitted HITs. Thus, completing many poor-paying but short HITs is a good tactic to remain or regain a high approval rate. Accordingly, the motivation to rapidly push one's approval rate should be less relevant for tasks that take several minutes. PFP is probably more important when crowdworkers perceive their input more valuable as identified in other contexts (Franke et al., 2013).

2.2.2 Consequences of Perceived Fairness in Pay for Crowdwork

Before we address the more tangible consequences, we want to underline ethical considerations. Some crowdworkers consider the work on MTurk as their primary source of income and they need the money to shop for daily needs incl. groceries (Ipeirotis, 2010; Silberman et al., 2010; Brawley and Pury, 2016). Furthermore, poor payments can make crowdworkers feel marginalized (Deng et al., 2016) and even with a fair wage crowdworkers are generally much less expensive than other methods of data generation (Berinsky et al., 2012; Lowry et al., 2016). Thus, we argue that ethical considerations alone should be enough reason to pay a fair wage (see Deng et al., 2016 for detailed ethical considerations and guidelines).

In addition to the ethical considerations, there are at least three consequences job requesters should consider when they decide whether to pay a fair wage or not. The first consequence might be the effort that crowdworkers invest into a task. For effort-based tasks, a lower PFP may lead to less effort (Ho et al., 2015; Ye et al., 2017). There seems to be no similar effect for attention-based tasks (Mason and Watts, 2009; Buhrmester et al., 2011; Goodman et al., 2013; Litman et al., 2015).

Second, job requesters should be aware that crowdworkers share information (Brawley and Pury, 2016). Paying a low amount of money makes it much more likely that crowdworkers engage in negative electronic word-of-mouth (eWOM) which leads to a bad reputation among crowdworkers. This communication takes place not just via little known individual blogs but via well-known platforms like the above mentioned Turkopticon. This makes it harder to find good participants for future tasks and it takes significantly longer to recruit participants in future because crowdworkers consider a job requesters' reputation (Benson et al., 2015; Silberman and Irani, 2016). This may also harm the reputation of similar job requesters (e.g., researchers). Stewart et al. (2015) estimate in a research context with a capture-recapture analysis that an average laboratory sample only accesses a pool of ~7,300 crowdworkers on MTurk and not the 500,000 crowdworkers that are advertised by Amazon. Furthermore, Kees et al. (2017) indicate that the average MTurk crowdworker completes around 160 surveys per month. This is alarming because “professional” survey takers are more likely to respond to a study (Chandler et al., 2014) and their probable lack of naivety can reduce effect size for many research topics (Chandler et al., 2015). Stewart et al. (2017) call it a tragedy of the commons because researchers access the same relatively small pool of participants and their surveys/experiments affect other researchers. Thus, researchers start searching for alternative platforms with more naïve participants (Peer et al., 2017) which mitigates but does not solve the problem. More and new crowdworkers participating in research on MTurk would diminish the problem of non-naivety because a larger pool of participants makes it more likely that the sample is not dominated by professional survey takers (Chandler et al., 2015). Difficulty of recruiting (naïve) crowdworkers constitutes a second possible consequence from low PFP.

Low PFP also leads to a decrease in overall crowdworker satisfaction (Ye et al., 2017). This is probably a major cause for the relatively high rate of crowdworker attrition on MTurk. Stewart et al. (2015) estimate that within three months around 26% of crowdworkers leave the pool of participants and get replenished by new crowdworkers. Higher pay rates should diminish crowdworker attrition and attract more crowdworkers to MTurk. However, decent pay rates offered by only a few job requesters, e.g., researchers, do not attract significantly more crowdworkers (Stewart et al., 2015). As a third consequence, MTurk may become a “market of lemons” (Akerlof, 1970) where high quality crowdworkers either leave MTurk or refuse to work for most job requesters. Then, the remaining job requesters are left with only bad crowdworkers (Chandler et al., 2013).

In summary, low PFP may lead to low effort, difficulty of future recruitment, and lower crowdworker satisfaction that leads to platform or even crowdwork exit. All these issues can lead to a lower quality of crowdwork.

3 Research Model and Hypotheses Development

We want to measure the crowdworkers’ pay expectations before executing a HIT because they may not even want to start it if they do not expect to be compensated fairly. We call the construct *perception of fairness in pay before task execution* (PFPbte). Crowdworkers who execute the HIT because of a high PFPbte (or despite of it) may change their *perception of fairness in pay after task execution* (PFPate) which we will consequently measure after HIT execution. A reason to execute a HIT despite low PFPbte may be the expectation that the execution will be pleasing to the crowdworker (in terms of joy, passing the time, or as a challenge) as discussed in literature (e.g., Buhrmester et al., 2011). As described in 2.2.1, there may be even other reasons for task execution but we will not consider them as they usually only occur in special cases (e.g., pay of 1 cent for a small HIT). We develop a model of PFPbte formation to answer the first research question (RQ 1).

We refer to organizational justice theory as the theoretical framework since it has been used for the analysis of PFP in many other work contexts. For example, it was applied to study employment selection (Gilliland, 1993) or pay and promotion decisions (Colquitt, 2001). In these contexts a relatively long-term relationship is established during which employers and workers have many rights and possibilities to act. For example, a worker who perceives his pay raise to be too low can reduce the quality of his work in the future without announcing it to the employer. In some countries, a disgruntled worker can appeal to the work council for help if he thinks the actions of the employer are unfair. In our case, the relationship may last only a few seconds or minutes and the crowdworker has almost no rights. If a job requester on MTurk does not pay for a HIT, for example, the crowdworker can only stop working for the job requester and place a bad review of the job requester (outside of MTurk). PFP can be influenced by two main factors: distributive fairness and procedural fairness (e.g., Gilliland, 1993). Distributive fairness concerns the perceived fairness of outcome distribution among actors while procedural fairness concerns the perceived fairness in the process of distributing that value.

Individuals often assess distributive fairness based on the equity rule (Adams, 1965) where they compare their work input and output to other involved parties. The job requester receives a return from work done by all crowdworkers and from further work executed by his firm. He keeps a part of it and distributes another part to all crowdworkers. It is almost impossible for a crowdworker to judge the total return to the job requester based on his contribution unlike in some situations of open innovation (see Franke et al., 2013 for an example referring to Threadless.com). Therefore, we assume that crowdworkers judge distribution fairness on the basis of comparisons with other crowdworkers. They all receive the same amount in the case of most HITs (exceptions occur when some crowdworkers are paid a bonus). Injustice can only occur in these cases if their "input" is not the same (e.g., in terms of time spent on the HIT or the quality of their work). They cannot really observe it, unless some minimum requirements are set by the job requester, but they are definitely aware of possible differences. Two characteristics of crowdworkers may lead to differences in execution time and work quality: education and experience with work on MTurk. We assume that well educated and highly experienced crowdworkers are less satisfied with a given (realistic) payment. By *realistic* payment, we refer to payments usually offered on MTurk. Of course, a very high payment would probably make almost everybody happy. Based on the equity rule, the more educated crowdworker may think that he delivers better work than less educated crowdworkers and feel that he deserves a higher payment (share) than the given payment. In other words:

H1: *The more educated a crowdworker the lower is PFPbte for a given payment.*

Experience with work on MTurk is measured by the number of executed HITs on MTurk and approval rate. However, just a few executed HITs can lead to a high approval rate and due to the high bar usually set for approval rate by job requesters (95%), there is little variation in approval rates of crowdworkers who are admitted to perform microtasks. The crowdworker who completed many HITs (with a high approval rate) may think that he delivers better work than less experienced crowdworkers and feel, following the equity rule, that he deserves a higher payment (share) than the given payment. In terms of PFPbte:

H2: *The more experienced a crowdworker the lower is PFPbte for a given payment.*

One aspect of procedural justice relates to the interaction between an employer and a worker (Bies and Moag, 1986). This is also referred to as interpersonal treatment (Gilliland, 1993). In microtasks, there is implicit communication, e.g., when a job requester issues a payment for a microtask. Explicit communication may occur when task instructions are unclear or work has not been accepted (Deng et al., 2016). This is how a crowdworker gains experience with a job requester. In case of MTurk, the experience of other crowdworkers can also be looked up on a website like Turkopticon. Therefore, this aspect of procedural justice can be operationalized by the variable "experience with job requester" which relates to past interactions with a job requester. We hypothesize:

H3: *Positive previous experience with a job requester leads to higher PFPbte than in the case of new (and otherwise unknown) job requesters for a given payment.*

The task description can be considered the informational aspect of procedural justice (Colquitt, 2001) since a good and clear description correctly informs crowdworkers about what needs to be done (Sheehan, 2017). Examples of descriptions are: "Transcribe all of the item names, UPCs and amounts from a grocery receipt" or "Locate objects in an image." The completion time is sometimes estimated by the job requester and specified next to the task description. The completion time depends, for example, on the number of questions to be answered, items to be searched, or similar, but also on the necessary cognitive effort. The difference of tasks in difficulty is often reflected in the pay but not always (Hirth et al., 2011). We assume that task complexity is inferred from the description (perhaps in combination with estimated completion time) and assume that it will influence PFPbte:

H4: *Tasks whose execution appears to be complex lead to lower PFPbte.*

It is possible that higher educated or more experienced crowdworkers will be less sensitive to HIT complexity than the lower educated or less experienced crowdworkers but we expect that the direction of influence will still be the same. Other aspects of procedural justice exist (e.g., Colquitt et al., 2001)

but in our case they are often not observable before executing a HIT. For example, very few HITs on MTurk describe how a job requester will decide about the acceptance of HITs.
We intend to test our model (see model 1 in Figure 1) in a field experiment (HIT 1) where we identify the participants' characteristics and their perceptions of a described HIT to measure their PFPbte. Then, we intend to let the participants execute the HIT (HIT 2) that we only described in the first HIT if they wish to do it. This will allow us to answer the research question about the effect of PFPbte on actual task execution (RQ 2). We assume that high PFPbte will lead to high participation. This can be explained by Self-Determination Theory (SDT; Ryan and Deci, 2000) which states that people can be extrinsically motivated by money to act (here: HIT execution). Since most crowdworkers participate for money (e.g., Paolacci et al., 2010; Silberman et al., 2010; Behrend et al., 2011; Kaufmann et al., 2011; Martin et al., 2013; Ye et al., 2017), we hypothesize:

H5: *High PFPbte leads to HIT execution.*

A similar assumption can be made for expectation of immaterial satisfaction (e.g., enjoyment). SDT posits that people can also be intrinsically motivated to act if the act itself is inherently enjoyable or interesting (Ryan and Deci, 2000). We measure the construct "joy expectation" already in HIT 1 and assume that the expectation is formed on the basis of the task description. Literature specifically suggests that some crowdworkers are partly intrinsically motivated and complete tasks for enjoyment because tasks may be fun or a good way to kill time (Ipeirotis, 2010; Buhrmester et al., 2011; Mason and Watts, 2009).

The corresponding hypothesis is:

H6: *High joy expectation leads to HIT execution.*

After forming PFP and joy expectations, crowdworkers decide whether to execute the HIT (HIT 2) that has previously been described (model 2). If they decide to perform the HIT they experience the actual cognitive effort and the actual joy from executing the task. This may change their PFP. Therefore, we measure it again, but call it PFPate this time. Crowdworkers on MTurk have, on average, a great amount of experience with common HITs like surveys, tagging pictures, or searching for information on the internet (e.g., Rand et al., 2014; Chandler et al., 2014; Kees et al., 2017; Hirth et al., 2011). This suggests that most crowdworkers should be able to accurately assess PFPbte for common tasks. Hence, assuming that task descriptions are accurate and well understood, we hypothesize that:

H7: *There is no significant difference between PFPbte and PFPate.*

The test of this hypothesis will help us to answer research question number 3 (RQ 3).

PFPate may have a number of consequences as described in 2.2.2. A crowdworker may decide to draw them just for himself or evaluate the HIT and the job requester publicly on a platform like Turkopticon. In this research, we concentrate on antecedents of PFP, the consequence of this assessment for task execution, and possible differences between PFP when assessed ex ante and ex post. Here, we do not model consequences of PFPate. All constructs and models are depicted in Figure 1.

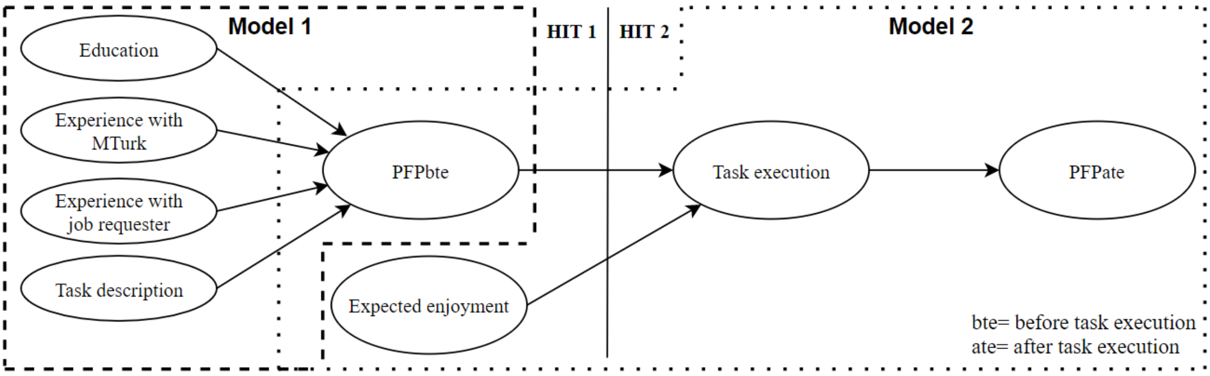

Figure 1. Research Model

4 Design of the Experiment

Figure 1 already indicates the main architecture of the experiment. We have conducted two surveys on MTurk with the same population of crowdworkers. We built up credibility with these crowdworkers as prompt payers with very few rejections. Therefore, we can approach them to form a group of crowdworkers who had a positive experience with us as a job requester. They performed the previous surveys with care and they answered personal questions (e.g., about their age and education) consistently across the two studies. In addition, they passed multiple attention checks. We will create another group of crowdworkers who never worked for us before. Therefore, we will have two groups (good experience and no experience with job requester). We will survey both groups about the number of HITs executed on MTurk and their education. The quantitative experience with MTurk is recorded on the platform so crowdworkers can easily retrieve their number of completed HITs. Note, if the number of completed jobs is a HIT requirement, the platform checks this automatically but job requesters, in general, have no direct and easy way to retrieve this characteristic of a crowdworker. Therefore, we simply ask our participants to state their total completed HITs. It is possible that crowdworkers may misrepresent this fact in a survey to be eligible for future HITs (Chandler and Paolacci, 2017; Wessling et al., 2017). Thus, we will clearly state that it is not a screening question which should decrease the economic incentive to lie. Further consistency checks are possible. We utilize Turkprime.com as an additional tool because it adds many useful features to MTurk (see Litman et al., 2017 for an overview). This tool allows us to effectively manage our HITs, HIT requirements, payments/rejections and crowdworkers.

We will randomly offer the crowdworkers two types of tasks that require different cognitive effort to which we can refer as simple task and complex task. An analysis of task categories and corresponding average payments reveals significant differences (Hirth et al., 2011). Simple tasks like searching, clicking, voting, or categorizing usually promise small payments while writing product reviews, writing an article, or software download and installation lead to relatively high payments. The description will describe the work as accurately as possible but within usual description lengths on MTurk. Estimated HIT length in minutes will be given, too. The estimation has been derived from execution by independent subjects (not authors) in a controlled experiment. The payment for HIT 2 will be disclosed to each crowdworker of HIT 1 according to the randomly offered task.

PFPbte will consist of slightly adapted items from (Ye et al., 2017). Joy expectation will be assessed as suggested by (Schulze et al., 2012) for their construct *enjoyment fit*. Crowdworkers will be paid for HIT 1 independently of their decision to perform HIT 2. If they decide to perform the second HIT, they get the payment promised in the first HIT. PFPate can be assessed in the same way as by (Ye et al., 2017). We plan to analyse Model 1 with partial least squares path modelling and the corresponding recommended quality criteria regarding the measurement models and the structural model (PLS-SEM; Hair et al., 2017). We will analyse Model 2 with logistic regression because PLS-SEM is not recommended for dependent binary/categorical variables (Hair et al., 2012). Appropriate tests of goodness of fit of the regression model and its parameters will be conducted to test its quality. The comparison of PFPbte and PFPate can be conducted with a simple paired t-test. We control for several crowdworker characteristics like gender, age, employment status and dependency on income from MTurk.

5 Summary

First, we have described problems in assessing PFP only ex post. Then, we have developed a model of PFPbte formation based on organisational justice theory adjusted to microtask crowdworking. Another model measures the impact of PFPbte on actual task execution. Some preliminary work for testing the models with real crowdworkers on MTurk has been done, namely a formation of a panel of crowdworkers who had good experience with us as a job requester. In addition, we designed a simple and a complex task, which we tested regarding comprehension and completion time. In the next steps, we will conduct the experiment that includes two HITs and examine the hypotheses we have posited in this report of our research in progress.

References

Adams, J. S. (1965). “Inequity in social exchange.” In: *Advances in experimental social psychology* 2. Ed. by L. Berkowitz. New York: Academic Press, 267-299.

Afuah, A., and Tucci C.L. (2012). “Crowdsourcing as a solution to distant search.” *Academy of Management Review* 37 (3), 355–375.

Akerlof, G. A. (1970). “The market for “lemons”: quality uncertainty and the market mechanism.” *Quarterly Journal of Economics* 84 (3), 488–500.

Behrend, T. S., Sharek, D. J., Meade, A. W., and Wiebe, E. N. (2011). “The viability of crowdsourcing for survey research.” *Behavior Research Methods* 43 (3), 800-813.

Belleflamme, P., Lamber, T., and Schwienbacher, A. (2014). “Crowdfunding: Tapping the right crowd.” *Journal of Business Venturing* 29 (5), 585-609.

Benson, A., Sojourner, A., and Umyarov, A. (2015). Can Reputation Discipline the Gig Economy? *Forschungsinstitut zur Zukunft der Arbeit Discussion Paper No. 9501.* URL: http://ftp.iza.org/dp9501.pdf (visited on 11/25/2017).

Berinsky, A. J., Huber, G. A., and Lenz, G. S. (2012). “Evaluating Online Labor Markets for Experimental Research. Amazon.com's Mechanical Turk.” *Political Analytics* 20 (3), 351–368.

Bies, R. J., and Moag, J. F. (1986). “Interactional justice: Communication criteria of fairness.” In: *Research on negotiations in organizations* 1. Ed. by R.J. Lewicki, B. H. Sheppard, and M. H. Bazerman. Greenwich, CT: JAI Press, 43-55.

Boudreau, K.J., Lacetera, N., and Lakhani, K. R. (2011). “Incentives and Problem Uncertainty in Innovation Contests: An Empirical Analysis.” *Management Science* 57 (5), 843.

Brawley, A. M., and Pury, C. L. (2016). “Work experiences on MTurk: Job satisfaction, turnover, and information sharing.” *Computers in Human Behavior* 54, 531-546.

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, 3-5.

Casey, L. S., Chandler, J., Levine, A. S., Proctor, A., and Strolovitch, D. Z. (2017). “Intertemporal Differences Among MTurk Workers: Time-Based Sample Variations and Implications for Online Data Collection.” *SAGE Open* 7:2, 1-15.

Casler K., Bickel, L., and Hackett, E. (2013). “Separate but equal? A comparison of participants and data gathered via Amazon’s MTurk, social media, and face-to-face behavioral testing.“ *Computers in Human Behavior* 29, 2156-2160.

Chandler, J., Mueller, P., and Paolacci, G. (2014). “Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers.” *Behav. Res. Methods* 46, 112–130.

Chandler, J., and Paolacci, G. (2017). “Lie for a Dime: When Most Prescreening Responses Are Honest but Most Study Participants Are Impostors.” *Social Psychological and Personality Science* 8 (5), 500-508.

Chandler, J., Paolacci, G., Peer, E., Mueller, P., and Ratliff, K. A. (2015). “Using Nonnaive Participants Can Reduce Effect Sizes.” *Psychological Science* 26 (7), 1131-1139.

Chandler, J., Paolacci, G., and Mueller, P. (2013). “Risks and Rewards of Crowdsourcing Marketplaces.” In: *Handbook of Human Computation*. Ed. by Springer New York, 377-392.

Chandler, J., and Shapiro, D. (2016). “Conducting Clinical Research Using Crowdsourced Convenience Samples.” *Annual review of clinical psychology* 12, 53–81.

Cohen-Charash, Y., and Spector, P.E. (2001). “The role of justice in organizations: A meta-analysis.” *Organ. Behav. Human Decision Processes* 86 (2), 278–321.

Colquitt, J. A. (2001). “On the dimensionality of organizational justice. A construct validation of a measure.” *Journal of Applied Psychology* 86 (3), 386–400.

Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H. and Ng, K. Y. (2001). "Justice at the millenium. A meta-analytic review of 25 years of organizational justice research." *Journal of Applied Psychology* 86 (3), 425–445.

Daly, T. M., and Nataraajan, R. (2015). "Swapping bricks for clicks: Crowdsourcing longitudinal data on Amazon Turk." *Journal of Business Research* 68 (12), 2603–2609.

Deng, X., Joshi, K. D., and Galliers, R. D. (2016). "The Duality of Empowerment and Marginalization in Microtask Crowdsourcing: Giving Voice to the Less Powerful through Value Sensitive Design." *MIS Quarterly* 40 (2), 279–302.

Durward, D., Blohm, I., and Leimeister, J. M. (2016). "Crowd Work." *Business & Information Systems Engineering (BISE)* 58 (4), 1-6.

Folger, R., and Konovsky, M. A. (1989). "Effects of Procedural and Distributive Justice on Reactions to Pay Raise Decisions." *Academy of Management* 32 (1), 115-130.

Ford, R. C., Richard, B., and Ciuchta, M. P. (2015). "Crowdsourcing: A new way of employing non-employees?" *Business Horizons* 58 (4), 377–388.

Fort, K., Adda, G., and Cohen, K. B. (2011). "Amazon Mechanical Turk: Gold Mine or Coal Mine?" *Computational Linguistics* 37 (2), 413-420.

Franke, N., Keinz, P., and Klausberger, K. (2013). "Does This Sound Like a Fair Deal? Antecedents and Consequences of Fairness Expectations in the Individual's Decision to Participate in Firm Innovation." *Organization Science* 24 (5), 1495–1516.

Gilliland, S. W. (1993). "The perceived fairness of selection systems: An organizational justice perspective." *Acad. Management Rev.* 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.

Goodman, J. K., and Paolacci, G. (2017). "Crowdsourcing Consumer Research." *Journal of Consumer Research* 44 (1), 196–210.

Guidelines for Academic Researchers (2017). URL: http://wiki.wearedynamo.org/index.php?title=Guidelines_for_Academic_Requesters (visited on 11/25/2017).

Hair, J. F., Hult, G. T. M., Ringle, C. M., and Sarstedt, M. (2017). "A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)." 2nd Ed., Sage Publications Ltd.

Hair, J. F., Sarstedt, M., Ringle, C. M. and Mena, J. A. (2012). "An assessment of the use of partial least squares structural equation modeling in marketing research." *Journal of the Academic Marketing Science* 40 (3), 414–433.

Hirth, M., Hoßfeld, T., and Tran-Gia, P. (2011). "Anatomy of a Crowdsourcing Platform - Using the Example of Microworkers.com." In: *Proceedings of the Workshop on Future Internet and Next Generation Networks*, FINGNet, Seoul, Korea, June 2011.

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

Hosseini, M., Shahri, A., Phalp, K., Taylor, J., and Ali, R. (2015). "Crowdsourcing: A taxonomy and systematic mapping study." *Computer Science Review* 17, 43-69.

Howe, J. (2006). "The Rise of Crowdsourcing." *Wired Magazine* 14.06, 1-5. URL: https://www.wired.com/2006/06/crowds/ (visited on 11/25/2017).

Ipeirotis, P. (2010). "Demographics of Mechanical Turk." *CeDER Working Papers*. New York University. URL: https://archive.nyu.edu/handle/2451/29585 (visited on 11/25/2017).

Kaufmann, N., Schulze, T., and Veit, D. (2011). "More than fun and money. Worker Motivation in Crowdsourcing - A Study on Mechanical Turk." In: *Americas Conference on Information Systems*, Detroit: Michigan, p. 1-11.

Kees J., Berry, C., Burton, S., and Sheehan, K. (2017). "An Analysis of Data Quality: Professional Panels, Student Subject Pools, and Amazon's Mechanical Turk." *Journal of Advertising* 46 (1), 141-155.

Litman, L., Robinson, J., and Abberbock, T. (2017). "TurkPrime.com. A versatile crowdsourcing data acquisition platform for the behavioral sciences." *Behavior research methods* 49 (2), 433–442.

Litman, L., Robinson, J., and Rosenzweig, C. (2015). "The relationship between motivation, monetary compensation, and data quality among US- and India-based workers on Mechanical Turk." *Behavior research methods* 47 (2), 519–528.

Lowry, P. B., D'Arcy, J., Hammer, B., and Moody, G. D. (2016). ""Cargo Cult" science in traditional organization and information systems survey research: A case for using nontraditional methods of data collection, including Mechanical Turk and online panels." *The Journal of Strategic Information Systems* 25 (3), 232–240.

Martin, D., Hanrahan, B. V., O'Neill, J., and Gupta, N. (2013). "Being A Turker." In: *17th ACM Conference on Computer Supported Cooperative Work and Social Computing*, Baltimore, MD.

Mason, W., and Suri, S. (2012). "A Guide to Conducting Behavioral Research on Amazon's Mechanical Turk." *Behavior Research Methods* 44 (1), 1-23.

Mason, W., and Watts, D. J. (2009). "Financial incentives and the 'performance of crowds'." In: *Proceedings of the ACM SIGKDD Workshop on Human Computation - HCOMP '09*, Paris: France. URL: https://pdfs.semanticscholar.org/1ae6/228dccf569d6990c7afc31282c40f9da23bc.pdf (visited on 11/25/2017).

Paolacci, G., Chandler, J., and Ipeirotis, P. G. (2010). "Running experiments on Amazon Mechanical Turk." *Judgment and Decision Making* 5 (5), 411-419.

Peer, E., Brandimarte, L., Samat, S., and Acquisti, A. (2017). "Beyond the Turk: Alternative platforms for crowdsourcing behavioural research." *Journal of Experimental Social Psychology* 70, 153-163.

Ramaswami, S. N., and Singh, J. (2003). "Antecedents and Consequences of Merit Pay Fairness for Industrial Salespeople." *Journal of Marketing* 67 (4), 46-66.

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., and Greene, J. D. (2014). "Social heuristics shape intuitive cooperation." *Nature Communications* 5 (3677), 1-12.

Ryan, R. M., and Deci, E. L. (2000). "Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions." *Contemporary Educational Psychology* 25, 54-67.

Sheehan, K. B. (2017). "Crowdsourcing research: Data collection with Amazon's Mechanical Turk." *Communication Monographs* 14 (6), 1–17.

Schmidt, G. B., and Jettinghoff, W. M. (2016). "Using Amazon Mechanical Turk and other compensated crowdsourcing sites." *Business Horizons* 59 (4), 391–400.

Schulze, T., Krug, S., and Schader, M. (2012). "Workers' task choice in crowdsourcing and human computation markets." In: *33rd International Conference on Information Systems*, Orlando: FL, 1-11.

Silberman, M. S., and Irani, L. (2016). "Operating an Employer Reputation System: Lessons from Turkopticon, 2008-2015." *Comparative Labor Law & Policy Journal* 37 (3), 505–542.

Silberman, M. S., Irani, L., and Ross, J. (2010). "Ethics and tactics of professional crowdwork." *XRDS*, 17:39–43, 39-43.

Steelman, Z. R., Hammer, B. I., and Limayem, M. (2014). "Data Collection in the Digital Age: Innovative Alternatives to Student Samples." *MIS Quarterly* 38 (2), 355–378.

Stewart, N., Chandler, J., and Paolacci, G. (2017). "Crowdsourcing Samples in Cognitive Science." *Trends in Cognitive Sciences* 21 (10), 736 – 748.

Stewart, N., Ungemach, C., Harris, A. J. L., Bartels, D. M., Newell, B. R., Paolacci, G., and Chandler, J. (2015). "The average laboratory samples a population of 7,300 Amazon Mechanical Turk workers." *Judgment and Decision Making* 10, 479-491.

Wessling, K. S., Huber, J., and Netzer, O. (2017). "MTurk Character Misrepresentation: Assessment and Solutions." *Journal of Consumer Research* 44, 211-230.

Ye, T., You, S., and Robert Jr., L. P. (2017). "When Does More Money Work? Examining the Role of Perceived Fairness in Pay on the Performance Quality of Crowdworkers." In: *Proceedings of the Eleventh International AAAI Conference on Web and Social Media*, University of Michigan.

Ye, H., and Kankanhalli, A. (2017). "Solvers' participation in crowdsourcing platforms: Examining the impacts of trust, and benefit and cost factors." *Journal of Strategic Information Systems* 26 (2), 101-117.