Leveraging Upfront Payments to Curb Employee Misbehavior: Evidence from a Natural Field Experiment*

John A. List†, Fatemeh Momeni‡

Tuesday 19th February, 2019

Abstract

We use a natural field experiment in which we hired over 2,000 workers from an online labor market to explore how upfront payment affects worker motivation and misbehavior on the job. We start with a simple theory that shows paying upfront can increase misbehavior through reducing the perceived costs of cheating, but it can decrease misbehavior through generating a gift-exchange effect. Motivated by the theory, we designed a task that provided workers with opportunities to reciprocate or misbehave. A unique aspect of our design is that we are permitted an opportunity to measure the curvature of the gift-exchange value of the upfront payment. Our results suggest paying workers upfront induces a gift-exchange effect that is *concave* in the share of total wage paid upfront. Moreover, the impact is strong enough to suggest that small upfront payments are a cost-effective means for an employer to curb employee misbehavior.

*We thank Joseph Seidel for excellent research assistance.
†Department of Economics, University of Chicago, Chicago, IL, USA. E-mail: jlist@uchicago.edu
‡Department of Economics, University of Chicago, Chicago, IL, USA. E-mail: fmomeni@uchicago.edu

1 Introduction

Employee misbehavior is a time worn problem organizations face daily.[1] Companies lose about 5% of their annual revenues because of various forms of internal misconduct (Association of Certified Fraud Examiners, 2016). For example, employees steal roughly $200 billion from US firms annually to supplement their income (Pierce, Snow and McAfee, 2015). One estimate suggests that in the retail industry alone, every year one in every 38 employees is apprehended for employee theft, costing firms $15 billion a year (NRF, 2015).[2] These facts, of course, are not lost on employers, who use various tools to reduce worker misbehavior on the job.

The most commonly used strategy to reduce misbehavior is monitoring and control. Companies pour an increasing amount of their resources into monitoring workers' activities. For instance, a study by American Management Association reported that in 2005: 76% of the surveyed employers monitored workers' website connections, 50% monitored employees' computer files, 55% monitored employees' emails, 51% tracked the amount of time employees spent on the phone and the numbers they called, and over 51% used video monitoring to counter theft and sabotage (AMA/ePolicy Institute Research, 2005). While the $200 million industry of employee monitoring has been estimated to grow to $500 million within the next four years,[3] research has shown a dark side of surveillance and control. Monitoring employees' behavior not only is costly, it can be counterproductive. Surveillance and control can crowd-out workers' intrinsic motivation and result in less trustworthy behavior and lower performance through signaling employer's distrust and inducing negative reciprocity (Frey 1993; Fehr and List 2004; Falk and Kosfeld 2006).

In this paper, we employ a natural field experiment to explore whether elements of the wage contract itself can be used to attenuate employee misbehaviors. Unlike recent work exploring the effect of bonuses on worker sabotage (Flory et al., 2016), our focus relates to one particular aspect of the wage contract: paying workers upfront. We start with a simple framework, which illustrates how paying a portion of the wage upfront influences the decision to misbehave through two channels. On the one hand, upfront payments can reduce misbehavior through signaling the employer's trust. Research suggests that contracts and work environments that signal trust can increase trustworthiness and favorable behaviors in workers through inducing positive reciprocity (Fehr and List 2004; Falk and Kosfeld 2006). Alternatively, paying the wage upfront reduces the perceived cost of misbehavior, especially shirking. A worker who fears the loss of remaining payments in the case of getting caught

[1] We adopt Kish-Gephart et al., (2010) definition of employee misbehavior: "Behavior that violates widely accepted moral norms such as theft, sabotage, lying to customers and misrepresentation of financial reports ... Other negative, counterproductive or deviant workplace behavior such as lateness, are not included because they do not necessarily violate widely accepted moral norms".

[2] http://hayesinternational.com/news/annual-retail-theft-survey/

[3] https://www.bostonglobe.com/business/2016/02/18/firms-step-monitoring-employee-activities-work/2l5hoCjsEZWA0bp10BzPrN/story.html

shirking would have less to lose if he has already received a portion of the wage upfront. Therefore, higher upfront payment reduces the opportunity cost to a worker from misbehaving. With the two forces acting in opposite directions, the overall effect of upfront payment on misbehavior remains ambiguous theoretically without more structure placed on the problem.

We formalize our predictions through a model in which a worker who exhibits gift-exchange motives towards the employer misbehaves less, if she perceives the employer to be "kinder".[4] Beyond understanding how upfront payment affects worker misbehavior, our model coupled with data from the experiment enable us to provide insight into the functional form of the gift-exchange value of upfront payment in worker's utility. Understanding the the functional form serves as an important step in designing an optimal contract.

To provide empirical content to our theory, we hired over 2,000 workers through an online labor market platform to perform a short task for payment. The task was to transcribe 10 images. Before starting to work on an image, a worker was required to report whether the image was readable. If the image was reported as readable, the worker was expected to transcribe it to text. If it was reported as not readable, then transcription was not necessary to receive payment. We randomized workers into four contracts across which we varied the upfront payment from 0% to 10%, 50% and 90% of the total pay. Workers received the remaining of their payments once they finished the task. We measure misbehavior in two ways: i) workers lying about whether the image was readable, and ii) accepting the contract and being paid some money upfront, but not completing the task. To explore how upfront payment affects worker misbehavior, we compare misbehavior on both the extensive and the intensive margins as we vary the share of upfront payment across treatments.

Our field experimental data provides three unique insights. First, the upfront payment curbs misbehaviors. Compared to the baseline, every treatment that has some upfront payment decreases misbehavior. For example, paying 10% of the payment upfront significantly decreases the propensity for individuals to misbehave: whereas 0.29 of workers misbehave in baseline, 0.21 misbehave when receiving 10% upfront. Second, we find a U-shaped relationship between misbehavior and upfront payment. As we increase the upfront payment to 50% and 90% of the total wage, the propensity to misbehave increases to 0.23 and 0.28. Interestingly, the average intensity of misbehavior per worker also exhibits a U-shaped relationship with the level of upfront payment. Through the lens of our model, we show that the U-shaped relationship between misbehavior and upfront payment implies that the gift-exchange effect is concave in the share of total wage paid upfront. This concavity suggests that while a small upfront payment signals the employer's trust and reduces misbehavior through inducing positive reciprocity, larger upfront payments may be less effective in reducing misbehavior as the fall in perceived cost of cheating offsets the gift-exchange effect.

[4]Previous research has shown that financial and non-financial incentives can both induce gift-exchange in workers (Bradler, et al., 2016; Dur, 2009; Kosfeld and Neckermann, 2011; Kube, Maréchal, and Puppe, 2012).

Our third insight is that while our findings suggest that paying workers a small portion of the total payment upfront reduces misbehavior, upfront payments come at a cost to the employer. To evaluate the cost-effectiveness of upfront payments, one should take into account not only the cost of misbehavior, but also the losses due to forgone upfront payments to workers who take the upfront money but do not complete the job. We provide insights into the cost-effectiveness of paying workers upfront by comparing the cost that the employer incurs for obtaining the same level of output produced, across our four treatments. We show that provided misbehavior is costly enough to the employer, paying 10% of the total wage upfront is more cost-effective than paying the entire wage after the work is completed. This result suggests that the employer can reduce misbehavior and save on labor costs simultaneously, through making a small share of the payment upfront. We show that the strategy of paying a large portion (90%) of the payment upfront is always dominated by the other three payment schemes such that regardless of how costly misbehavior is to the employer, it is never optimal to pay a large portion of the wage upfront.

Our work shares similarities with several strands of behavioral research. One literature that we view as a cousin to our approach is the work that uses "clawback" incentives to induce behavioral change (e.g. Hossain and List 2012, Fryer et al. 2012, Levitt et al. 2016, and Bulte et al. 2018, Imas et al., 2016). The main innovation in this literature is that agents receive an upfront payment which they have to return in case their productivity or output fails to meet a certain threshold (i.e., a bonus scheme with upfront payments). This work shows that an upfront bonus with clawback can induce higher productivity amongst workers. In this literature, the upfront bonus leverages loss-aversion to induce greater effort, though the designs do not always rule out other channels. In our study, we use the upfront payment to leverage gift-exchange but can rule out other effects, such as loss aversion, by design.

Another branch of related work is the multi-tasking literature, which draws inspiration from the seminal principal-agent theories of Holmstrom and Milgrom (1991) and Baker (1992). The theoretical work is complemented by several empirical studies exploring aspects of the principal agent theory (e.g. Lazear, 2000; Paarsch and Shearer, 2000; Shearer, 2004; Al-Ubaydli et al., 2015; Hong et al., 2018; Bulte et al., 2019). Our study contributes to this research in that we find incentives operate effectively in the sense that they increase productivity along the contracted dimension. But, workers do trade off quality for quantity when quality is not contracted. Similar to Hong et al. (2018), these findings are in line with the standard-principal agent theories.

The remainder of our paper is organized as follows. Section 2 presents a simple framework through which we formalize our predictions on how upfront payments affect misbehavior and what can be inferred about the curvature of the gift-exchange value of upfront payment from the relationship between misbehavior and the level of upfront payment. Section 3 presents the experimental design. Section 4 presents the findings. Section 5 discusses the cost-effectiveness of paying upfront and section 6 concludes.

2 A Simple Framework

Consider a worker i in a job contract which offers a wage W. The contract offers a portion of the wage ($u_T \times W$) upfront and the remaining part of the wage ($(1 - u_T) \times W$) upon the completion of the job. The upfront money is paid before the worker starts the job and cannot be taken back to the principal. For simplicity, we assume that the worker makes only one decision: the level of misbehavior (or cheating), X, which can reflect various measures of harmful acts such as stealing from the employer, lying to the employer, and shirking on the task.[5] Thus, the employer's payoff is decreasing in X. The worker chooses X to maximize the following utility function:

$$U(X, W, u_T) = u_T \times W + (1 - u_T) \times W \times (1 - P(X)) + B(X) + A(u_T)(-X)$$

$$s.t. \quad X \geq 0$$

The first two terms represent the expected monetary payments to the worker. A worker who misbehaves would fear the loss of the remaining payments, in the case that she gets caught. We assume that the worker's perceived probability of getting caught P(X) is increasing in the intensity of cheating and that $P(0) = 0$. The worker engages in misbehavior because she derives a direct utility of $B(X)$. $B(X)$ may come in the form of either direct monetary benefit (e.g. stealing from the employer), or non-monetary benefit (e.g. lower cost of effort when misbehavior is shirking on the job). In the latter case, $B(X)$ can also be expressed as $-C(e)$ where e represents worker's effort. To keep the analysis tractable, we assume that B (C) is a concave (convex) and increasing function of misbehavior (effort):

$$B'(X) > 0; \quad B''(X) < 0$$ [6]

Finally, the last term in the utility function captures the gift-exchange effect of the upfront payment. If u_T serves as a signal of employer's trust, the worker may reciprocate this trust by lowering X. We assume that $A(u_T)$ is a monotonically increasing function of u_T. The first order condition to the worker's problem can be expressed as:

$$\frac{\partial U(X, W, u_T)}{\partial X} = -P'(X)W(1 - u_T) + B'(X) - A(u_T) = 0$$

[5] In this simplified model, worker's effort can be modeled as an inverse function of misbehavior, or cheating, such that higher effort corresponds to lower level of cheating.

[6] Or alternatively: $C'(e) > 0$ and $C''(e) > 0$.

Using the Implicit Function Theorem (IMT), we can derive the following condition, which tells us how the optimal level of misbehavior (X^*) responds to changes in the upfront payment u_T:

$$\frac{dX^*}{du_T} = -\frac{WP'(X^*) - A'(u_T)}{-P''(X^*)W(1-u_T) + B''(X^*)}$$

For simplicity, we assume that P is linear in X such that $P'(X) = b$ and $P''(X) = 0$. Therefore, we can simplify the above expression to: $\frac{dX^*}{du_T} = -\frac{Wb - A'(u_T)}{B''(X^*)}$. One can see that misbehavior increases with u_T if $Wb > A'(u_T)$ and it decreases with u_T if $Wb < A'(u_T)$.

Our experimental design, described below, therefore would allow the curvature of $A(u_T)$ to be inferred from the relationship between X^* and u_T. To see this point, let us define u^* to be the level of upfront payment for which $Wb = A'(u^*)$. If $A(u_T)$ is convex in u_T, cheating increases with u_T for all $u_T > u^*$ it and decreases with u_T or all $u_T < u^*$. Therefore, (as long as $u^* \notin \{0, 1\}$), X^* would exhibit an inversed-U-shaped relationship with u_T. Alternatively, an inversed-U-shaped X^* (u_T) would imply that $A(u_T)$ is a convex function of u_T. Similarly, one can show that a U-shaped $X^*(u_T)$ would imply that $A(u_T)$ is a concave function of u_T.[7]

3 Experimental Design

With our simple theory as a guide, we conducted a natural field experiment with workers from Amazon Mechanical Turk (MTurk), an online labor market platform where businesses and individuals recruit workers to complete short tasks for payment. We acted as an employer and posted a Human Intelligence Task (HIT) on MTurk, which provided some general information about the task and invited interested workers to our website where the task was to be carried out (see Appendix C). Interested subjects who landed on our website were provided with unique IDs, which they had to submit though the MTurk HIT before they were randomized into treatment. The unique IDs were used for payment purposes to link workers on our website to their MTurk account.

Contract. Upon landing on our website, workers were randomized into one of four treatments and were provided with treatment-specific contracts, which included information about the task and payment. The contract offered a total wage of \$1.20 to workers for completing the task.[8] Across the four treatments, which we denote as Baseline, Upfront10, Upfront50 and Upfront90, we varied the percentage of the total wage (from 0% to 90%: $u_T \in \{0\%, 10\%, 50\%, 90\%\}$) that we paid workers upfront upon accepting our contract . The remainder of the wage payment was made immediately

[7] Note that if $X^*(u_T)$ is monotonically increasing or monotonically decreasing in u_T ($u^* = \{0, 1\}$), we cannot learn the curvature of $A(u_T)$ from $X^*(u_T)$.

[8] The average hourly earnings on MTurk lies between \$1 and \$5 (Ross et al., 2010; Paolacci et al., 2010; Berg 2016). Given that an average worker spent about 15-20 minutes on our website, the average hourly earnings of our workers falls within this estimated range.

upon completion of the task. The treatment-specific parameter u_T was chosen such that it would allow us to detect non-monotonicities in misbehavior as a function of the level of upfront payment (u_T), which is critical for uncovering the curvature of $A(u_T)$. We explicitly asked workers to accept the contract, only if they intended to complete the task. Table 1 summarizes our experimental design.

Table 1: Treatments

	Baseline	Upfront10	Upfront50	Upfront90
Total Payment	\$1.20	\$1.20	\$1.20	\$1.20
Upfront Payment	\$0	\$0.12	\$0.60	\$1.08

Task. Each worker who accepted the contract was provided with 10 images of short scanned texts to transcribe. Before starting to work on an image, the worker was required to report whether the image was readable or not. If the worker reported an image as unreadable, then she skipped it and moved on to the next image. Workers were instructed that it was possible for some images to be illegible and reporting those as unreadable was not going to reduce their payment. By offering a fixed payment regardless of the number of images a worker actually transcribed, we provided workers with an opportunity to misreport perfectly readable images as unreadable and avoid transcribing them. We refer to such behavior as *type-1* cheating and use it as our first measure of misbehavior on the job. Note that *type-1* cheating involves explicit lying about the quality of an image as well as shirking on the job.

Besides lying about image legibility, a worker could also cheat by accepting the contract, receiving the upfront payment, and not completing the task. Since the upfront payment that was paid to a worker could not be returned, a worker could accept the contract and keep the upfront payment without transcribing images. Accepting the contract without completing the task serves as our second measure of misbehavior on the job. We refer to this behavior as *type-2* cheating. Type-2 cheating can be thought of as a less severe form of misbehavior, as it does not involve active lying, while it entails the dishonest act of accepting the contract and receiving the upfront money, without actually completing the task. The workers were required to complete the task within two hours of accepting the contract.

Payment. MTurk employers (also known as requesters) have two means for making payments to workers: reward and bonus. Each requester who posts a HIT is required to specify a "reward" for each assignment s/he posts, which is paid to the worker if (and as soon as) the requester approves the HIT submitted by the worker. In addition to the reward, the requester can also make multiple extra payments in the form of "bonuses". We set a reward of \$0.10 for our assignments, which was paid to all workers who landed on our website through the MTurk HIT, and submitted their unique IDs (which we used to link workers on our page to their MTurk account) before they were randomized into treatments. This \$0.10 reward was paid to a worker regardless of whether she accepted the contract. The wage, which consisted of an upfront payment and a remaining payment, was paid to workers through two sets of "bonuses". Paying the wages through bonuses (as opposed to reward) enabled us to split the wage payments to an *upfront* and a *remaining* part, which was essential for our design.[9] Note that once a payment is made to a worker, a requester on MTurk cannot take back the paid money from the worker. This feature enables us to rule out loss-aversion as a channel.

In summary, the timeline of our field experiment was as follows: 1) workers were recruited through MTurk (link to our webpage); 2) treatment-specific contracts were offered on our website 3) workers decided whether to accept the contract; 4) workers received the upfront payment if they accepted the contract; 5) workers received more detailed instructions on how to perform the task; 6) workers transcribed images; 7) workers received the remaining of their wage. With our design choice we are estimating the "total effect" of the upfront payment in the spirit of employment in the Gig Economy: we advertise a wage contract, people decide whether to opt in, and we only observe behavior of those who opt in. This type of design is unable to parse treatment from selection, but provides the total estimated effect cleanly. As we will highlight next, in practice there were only a few workers who did not accept our contract in any treatment so selection effects are minimal empirically.

4 Results

The field experiment commenced in October 2016 and we continued collecting data until we gathered at least 500 subjects in each treatment cell who accepted the contract.[10] The rate of accepting the contract lies between 96.3% to 97.1%, and is not significantly different across any pair of treatments, [11] suggesting that selection is not likely to play a major role as a potential driving force behind any observed differences in misbehavior across treatments. This leads to Result 1:

Result-1: The rates of accepting the contract are between 96.3% to 97.1% and are not significantly different across treatments.

To explore how upfront payment affects cheating on the job, we compare both the propensity

[9] Appendix D presents the contract.

[10] The share of cheaters in our pilot experiment with 10% upfront payment was slightly over 20%. Our power calculation revealed that (for power=0.80 and $\alpha = 0.05$) we would have needed around 450 subjects in each treatment in order to be able to detect about 30% change in the mean share of cheaters (List, Sadoff & Wagner, 2011).

[11] P-values from two-sample test of proportions with the null hypothesis of equal proportions are: $P_{Baseline-Upfront10} = 0.50$; $P_{Baseline-Upfront50} = 0.62$; $P_{Baseline-Upfront90} = 0.99$; $P_{Upfront10-Upfront50} = 0.86$; $P_{Upfront10-Upfront90} = 0.50$; $P_{Upfront50-Upfront90} = 0.61$

Table 2: Acceptance Rates

	Baseline	Upfront10	Upfront50	Upfront90
Acceptance Rate	96.4%	97.1%	96.9%	96.3%
# workers who accepted the contract	502	502	502	501

Notes: The percentage and the total number of workers who accepted the contracts are presented.

to cheat, as well as the average intensity of cheating by each worker, as we increase the upfront payment from 0% to 10%, 50% and 90% of the total wage.

4.1 Propensity to Cheat

We refer to a worker as a *type-1 cheater* if she engages in type-1 cheating and as a *type-2 cheater* if she engages in type-2 cheating. We refer to a worker as a *cheater* if she engages in either type-1 or type-2 cheating. Panel-a in Table 3 and Figure 1 illustrate how the share of cheaters change with the level of upfront payment. Both the table and figure reveal that as we increase the upfront payment from 0% to 10% of the total wage, the share of cheaters significantly drops, from 0.289 to 0.209 ($P_{Baseline-Upfront10} = 0.004$).[12] This result is consonant with Flory et al. (2016) in that the wage contract itself can be used as an important disciplining device in the workplace.

Interestingly, as we further increase the upfront payment from 10% to 50%, the share of cheaters increases to 0.227 but this increase is not significant at conventional levels ($P_{Upfront10-Upfront50} = 0.490$). By increasing the upfront payment from 50% to 90% of the total wage, the share of cheaters further increases to 0.279 ($P_{Upfront50-Upfront90} = 0.057$), a level that is not significantly different from baseline ($P_{Baseline-Upfront10} = 0.740$).

As the table and figure show, similar U-shaped patterns emerge when we explore the share of type-1 and type-2 cheaters, separately. The share of type-1 cheaters drops from 0.096 in the baseline to 0.056 in Upfront10 ($P_{Baseline-Upfront10} = 0.017$). A further increase of the upfront payment to 50% of the wage increases the share of type-1 cheaters to 0.068, but this increase is not significant at conventional levels ($P_{Upfront10-Upfront50} = 0.430$). Likewise, an increase of the upfront payment from 50% to 90% of the wage increases the share of type-1 cheaters to 0.104 ($P_{Upfront50-Upfront90} = 0.041$). Similarly, from the Baseline to Upfron10, the share of type-2 cheaters drops from 0.209 to 0.159 ($P_{Baseline-Upfront10} = 0.042$). The share of type-2 cheaters increases (but not significantly so) as the upfront pay further increases to 50% and 90% of the wage, to 0.161 and 0.184 ($P_{Upfront10-Upfront50} = 0.930$; $P_{Upfront50-Upfront90} = 0.350$). The U-shaped relationship between the share of cheaters and upfront payment is robust to restricting the data to only those workers who finished the task and received the full payment (of $1.20): the

[12] All p-values reported in this section are from two-sample test of proportions with the null of equal proportions.

Table 3: The Effect of Upfront Payment on Cheating

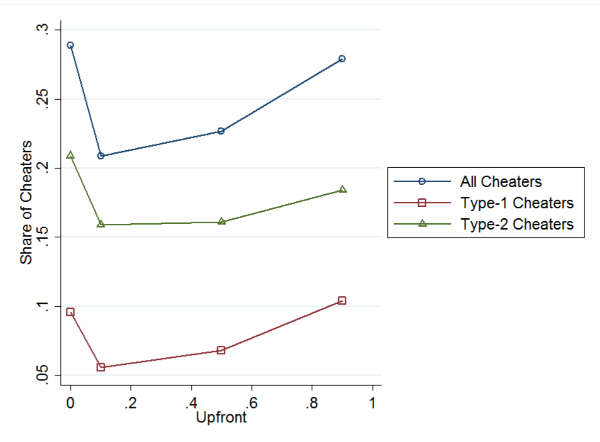
	Baseline	Upfront10	Upfront50	Upfront90
Panel-a: Propensity to Cheat				
	All workers			
Share of cheaters	0.289	0.209	0.227	0.279
	(0.020)	(0.018)	(0.019)	(0.020)
Share of Type-1 cheaters	0.096	0.056	0.068	0.104
	(0.013)	(0.010)	(0.011)	(0.014)
Share of Type-2 cheaters	0.209	0.159	0.161	0.184
	(0.018)	(0.016)	(0.016)	(0.017)
	Workers who accepted the contract			
Share of cheaters	0.101	0.059	0.078	0.117
	(0.015)	(0.012)	(0.013)	(0.016)
Panel-b: Intensity of Cheating				
	All workers			
Average cheating per worker	2.034	1.539	1.574	1.852
	(0.164)	(0.150)	(0.150)	(0.159)
Average type-1 cheating per worker	0.217	0.145	0.135	0.248
	(0.042)	(0.038)	(0.031)	(0.046)
Average type-2 cheating per worker	1.817	1.394	1.438	1.605
	(0.149)	(0.138)	(0.137)	(0.146)
	Workers who accepted the contract			
Average cheating per worker	0.254	0.159	0.159	0.276
	(0.052)	(0.044)	(0.036)	(0.054)

Notes: Panel-A presents the share of workers who cheated. Panel-B presents the average intensity of cheating per worker, defined as the number of images that a worker who accepted the contract either misreported as unreadable (type-1) or skipped transcribing by not completing the task (type-2). Standard errors are in parentheses.

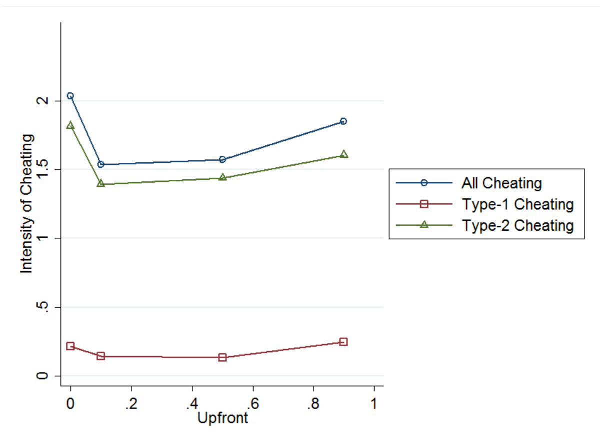
share of cheaters who finished submitting all 10 images drops from 0.101 in the baseline to 0.059 in Upfront10 ($P_{Baseline-Upfront10} = 0.028$). From Upfront10 to Upfront50, the share of cheaters who finished the task increases but not significantly so. Moving from Upfront50 to upfront90, the share of cheaters significantly increases from 0.078 to 0.117 ($P_{Upfront50-Upfront90} = 0.058$).

Taken together, a second result emerges:

Result-2: There is a U-shaped relationship between the propensity to cheat and the upfront payment: as the share of total payment made upfront increases from 0% to 10%, fewer workers cheat. As upfront money further increases from 10% to 50%, the share of cheaters increases, but not significantly. As we further increase the upfront from 50% to 90% of the total payment, the share of cheaters increases significantly, to a level which is not significantly different from the Baseline.



(a) Share of cheaters



(b) Average intensity of cheating

Figure 1: Cheating as a function of Upfront Payment

4.2 Intensity of Cheating

We measure the intensity of cheating by the number of images that a worker who accepted the contract either misreported as unreadable (type-1) or skipped transcribing by not completing the task (type-2). As upfront payment increases from 0% to 10% of the total pay, the average intensity of cheating per worker falls by roughly 25 percent: from 2.034 to 1.539 images per worker ($P_{Baseline-Upfront10} = 0.006$).[13] As the upfront payment further increases from 10% to 50%, the intensity of cheating slightly increases to 1.574, but this increase is not significant ($P_{Upfront10-Upfront50} = 0.55$). A further increase of upfront pay from 50% to 90% further increases

[13] All the p-values reported in analysis of the intensity of cheating are from Two-sample Wilcoxon rank-sum (Mann-Whitney) test, against the null hypothesis of equal means.

the intensity of cheating to 1.852 ($P_{Upfront50-Upfront90} = 0.080$).

A similar U-shaped pattern emerges when we study the intensity of type-1 or type-2 cheating, separately, as shown in panel-b of Figure 1 and Table 3. As the share of payment made upfront increases from 0% to 10%, the average number of readable images that a worker misreports as unreadable significantly decreases ($P_{Baseline-Upfront10} = 0.018$). An increase in the upfront payment from 10% to 50% does not affect type-1 cheating significantly, but a further increase from 50% to 90% significantly increases the average intensity of type-1 cheating per worker ($P_{Upfront50-Upfront90} = 0.040$). Similarly, from baseline to Upfront10, the average number of images that a worker who has accepted the contract skips transcribing by simply not completing the job significantly falls ($P_{Baseline-Upfront10} = 0.050$). As we further increase the upfront pay from 10% to 50% and from 50% to 90%, the intensity of type-2 cheating increases, but these increases are not significant ($P_{Upfront10-Upfront50} = 0.890$; $P_{Upfront50-Upfront90} = 0.379$).

The U-shaped relationship between the intensity of cheating and upfront payment is robust to restricting the sample to those workers who completed the task and thus received full payment. Increasing the upfront payment from 0% to 10% of the wage significantly reduces the average cheating per worker who finished the task from 0.254 to 0.159 ($P_{Baseline-Upfront10} = 0.028$). A further increase of the upfront payment to 50% does not affect the intensity of cheating by these workers ($P_{Upfront10-Upfront50} = 0.282$), but an additional increase to 90% of the total wage increases the average cheating by those who finished the task to 0.276 ($P_{Upfront50-Upfront90} = 0.058$).

The U-shaped relationship between cheating and upfront payment translates into an inversed U-shaped relationship between worker productivity (measured by the number of transcribed images) and the share of payment made upfront. The average number of images transcribed per worker significantly increases from 7.147 (std.err.=0.149) in the baseline to 7.580 (std.err.=0.138) in the Upfront10 treatment ($P_{Baseline-Upfront10} = 0.009$), but does not significantly change from Upfront10 to Upfront50. A further increase in the upfront payment from Upfront50 to Upfront90, significantly decreases the average productivity from 7.564 (std.err.=0.137) to 7.287 (std.err.=0.146) transcribed images per worker ($P_{Upfront50-Upfront90} = 0.053$), which is not significantly different from the Baseline ($P_{Baseline-Upfront90} = 0.75$).

In summary, our findings suggest that compared to the baseline of paying the entire wage after the job is completed, an employer can reduce worker misbehavior on the job by paying a small share of the total wage upfront. However, paying a large portion (90%) of the wage upfront is not effective in reducing misbehavior.[14] This leads to a third result:

Result-3: There is a U-shaped relationship between the average intensity of cheating per worker and the level of upfront payment. As the share of wage paid upfront increases from 0% to 10%, workers

[14] In Appendix A, we explore whether quality of transcription was affected by upfront payment and show that we do not find any significant relationship between the upfront payment and the quality of transcription.

cheat more intensely. From 10% to 50% upfront payment, the intensity of cheating per worker does not change significantly. As we further increase the upfront payment to 90% of the total wage, the intensity of cheating increases to a level close to the Baseline.

Through the lens of our model in Section 2, the U-Shaped relationship between cheating and the upfront payment implies that the gift-exchange value of upfront payment in worker's utility –A(u)– is a concave function of the share of payment made upfront. This suggests that a similar increase in the upfront payment induces a larger reciprocal effect in reducing worker misbehavior when upfront payment is small rather than large. The concavity of A(u) is consistent with the idea that the upfront payment serves as a signal of trust to workers and the marginal increase in signaling value of upfront payment takes its largest value around $u = 0$, and decreases as u increases, leading to Result 4:

Result-4: The U-shaped relationship between upfront payment and cheating implies that $A(u)$, which captures the gift-exchange value of upfront payment in worker's utility, has a concave relationship with u.

5 Discussion

Our findings suggest that paying a small portion of the total wage upfront reduces cheating on the job. However, upfront payments do not come at zero cost to the employer. Workers who receive payments upfront may take the money and not complete the task. Since payments that are made upfront cannot be returned to the employer, the low level of cheating in Upfront10 treatment does not necessarily mean that paying 10% of the wage upfront is also the most cost-effective scheme for the employer.

To investigate the cost-effectiveness of upfront payments, we compare an employer's cost of obtaining the same level of output (images transcribed) across different treatments. For simplicity, we assume that type-2 cheating is not costly and the employer only suffers a loss of b per instance of type-1 cheating.[15] The cost that a principal bears to get one image transcribed under each treatment can be expressed as the average labor cost divided by the average number of images transcribed per worker. The average worker cost in a treatment T, (C^T) is the sum of the wages paid per worker (C_1^T) and the losses due to instances of type-1 cheating per worker (C_2^T). We can express the average cost of a transcribed image in the treatment T as:

$$C^T(1) = \frac{C_1^T + C_2^T}{a^T} = \frac{p^T \times W + (1 - p^T) \times u^T \times W}{a^T} + \frac{b \times l^T}{a^T},$$

[15] This simplification does not affect our general conclusions. Given that type-2 cheating followed a similar pattern to type-1 cheating, taking into account the cost of type-2 cheating in our cost-benefit analysis will only make our conclusions stronger.

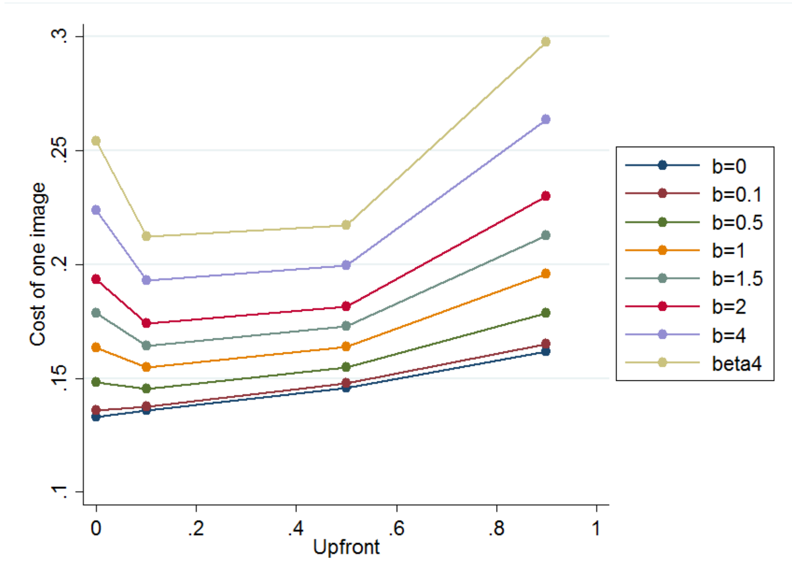


Figure 2: Employer's Cost of One Transcribed Image

where a^T represents the average number of images transcribed per worker; p^T is the probability of completing the task by each worker (which equals 1 minus the share of type-2 cheaters); W is the total wage; u^T is the portion of the total wage paid upfront; and l^T is the average level of type-1 cheating per worker.

Figure 2 provides an ocular depiction of the average cost to the principal for obtaining one transcribed image, for various values of b, as the upfront payment increases from 0% to 90% of the total wage. As shown in the Figure, while for small values of b (less than about \$0.20 per type-1 cheating), paying no upfront money is the most cost-effective payment structure among our four treatments, as b increases beyond \$0.20, paying 10% of the total wage upfront strictly dominates all other contracts.

In other words, if type-1 cheating is costly enough to the employer, s/he will benefit from paying a small portion of the total wage upfront, as workers who receive the upfront money cheat significantly less than those who do not. Note that regardless of the size of b, it is never optimal to pay 90% of the total wage upfront. If we take into account possible costs associated with type-2 cheating, our conclusions become even stronger. As type-2 cheating is also minimized in our Upfront10 treatment, if an employer also suffers from losses associated with type-2 cheating, Upfront10 dominates the Baseline for even smaller values of b.

6 Conclusions

Since the beginning of humankind principals have explored contracts that dictate agent actions. In a natural field experiment that allows both adverse selection and moral hazard in a Gig economy work setting, we provide empirical evidence on how upfront payments can be used to reduce worker misbehavior. We propose a simple model to provide insights on how paying a portion of wage upfront can mitigate cheating on the job, through signaling employer's trust and inducing positive reciprocity. We document a U-shaped relationship between cheating and the share of wage paid upfront. Compared to the baseline of no upfront payment, paying 10% of total wage upfront significantly reduces cheating. A further increase in upfront from 10% to 50% of the total wage does not have a significant impact on cheating, while increasing upfront from 50% to 90% significantly increases misbehavior on the job.

We evaluate the the cost-effectiveness of upfront payments by taking into account losses due to forgone upfront money to workers who did not complete the job. We show that as long as misbehavior is costly enough to the principal, paying a small (10%) share of the total wage upfront can be more cost-effective than paying the entire wage upon work completion. Yet, regardless of how costly cheating is to the principal, it is never cost-effective to pay a large (90%) share of the wage upfront. In addition to providing insights into the relationship between upfront payment and worker misbehavior, our experimental design allows us to learn about the curvature of the gift-exchange value of upfront payment in a worker's utility. Our simple theoretical analysis along with the finding of the U-shaped relationship between misbehavior and upfront payment suggest that the gift-exchange value of upfront payment is a concave function of the share of payment made upfront, such that the marginal increase in the gift-exchange effect decreases as the upfront payment increases.

References

[1] Al-Ubaydli, O., Gneezy, U., List, J. A., & Lee, M. S. (2015). "Carrots that look like Sticks: Toward an Understanding of Multitasking Incentive Schemes." *Southern Economic Journal*, 81(3), 538-561.

[2] AMA/ePolicy Institute Research. (2005). "Electronic Monitoring and Surveillance Survey ".

[3] Association of Certified Fraud Examiners. (2016). "Report To the Nations on Occupational Fraud and Abuse".

[4] Baker, G.P. (1992). "Incentive Contracts and Performance Measurement." *Journal of Political Economy*, 100(3), 598-614.

[5] Berg, J. (2016). "Income Security in the On-Demand Economy: Findings and Policy Lessons from a Survey of Crowdworkers." *Comparative Labor Law & Policy Journal,* 37(3).

[6] Bradler, C., Dur, R., Neckermann, S., & Non, A. (2016). "Employee Recognition and Performance: A Field Experiment." *Management Science*, 62(11), 3085-3099.

[7] Bulte, E.H., List, J.A., & van Soest, D.P. (2018), "Toward an Understanding of the Welfare Effects of Nudges: Evidence from a Field Experiment in Uganda." *Wageningen University: Working Paper.*

[8] Dur, R. (2009). "Gift Exchange in the Workplace: Money or Attention?" *Journal of the European Economic Association*, 7(2-3), 550-560.

[9] Falk, A., & Kosfeld, M. (2006). "The Hidden Cost of Control." *American Economic Review*, 96(5), 1611-1630.

[10] Fehr, E., & List, J. A. (2004). "The Hidden Costs and Returns of Incentives–Trust and Trustworthiness among CEOs." *Journal of the European Economic Association*, 2(5), 743-771.

[11] Flory, J., Leibbrandt, A. & List, J. A. (2016). "The Effects of Wage Contracts on Workplace Misbehaviors: Evidence from a Call Center Natural Field Experiment." *NBER Working Paper*, No. w22342.

[12] Hossain, T. & List, J.A. (2012). "The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations." *Management Science*, 58(12), 2151-2167.

[13] Frey, B. S. (1993). "Does Monitoring Increase Work Effort? The Rivalry with Trust and Loyalty." *Economic Inquiry*, 31, 663-670.

[14] Fryer Jr, R. G., Levitt, S. D., List, J. A., & Sadoff, S. (2012). "Enhancing the efficacy of teacher incentives through loss aversion: A field experiment." *NBER Working Paper*, No. w18237

[15] Holmstrom, B., & Milgrom, P. (1991). "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. " *Journal of Law, Economics and Organization,* 7, 24-52.

[16] Hong, F., Hossain, T., List, J. A., & Tanaka, M. (2018). "Testing the Theory of Multitasking: Evidence from a Natural Field Experiment in Chinese Factories ." *International Economic Review,* 59(2), 511-536.

[17] Imas, A., Sadoff, S., & Samek, A. (2016). "Do people anticipate loss aversion?." *Management Science*, 63(5), 1271-1284.

[18] Kish-Gephart, J. J., Harrison, D. a, & Trevio, L. K. (2010). "Bad apples, bad cases, and bad barrels: meta-analytic evidence about sources of unethical decisions at work." *The Journal of Applied Psychology*, 95(1), 1-31.

[19] Kosfeld, M., & Neckermann, S. (2011). "Getting more work for nothing? Symbolic awards and worker performance." *American Economic Journal: Microeconomics*, 3(3), 86-99.

[20] Kube, S., Maréchal, M. A., & Puppe, C. (2012), "The Currency of Reciprocity - Gift-Exchange in the Workplace." *The American Economic Review*, 102(4), 1644-1662.

[21] Lazear, E.P. (2000). "Performance Pay and Productivity." *American Economic Review*, 90(5), 1346-1361.

[22] Levitt, S., List, J. , Neckermann, S., & Sadoff, S. (2012). "The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance", *NBER Working Paper*, No. 18165.

[23] List, J. A., Sadoff, S., & Wagner, M. (2011). "So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design." *Experimental Economics,* 14(4), 439-457.

[24] Paarsch, H.J., & Shearer, B. (2000). "Piece rates, fixed wages, and incentive effects: Statistical evidence from payroll records." *International Economic Review*, 41(1), 59-92.

[25] Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). "Running experiments on Amazon Mechanical Turk." *Judgment and Decision Making,* 5(5).

[26] Pierce, L., Snow, D. C., & McAfee, A. (2015). "Cleaning House: The Impact of Information Technology Monitoring on Employee Theft and Productivity." *Management Science*, 61(10), 2299-2319.

[27] Ross, J., Irani, L., Silberman, M., Zaldivar, A., & Tomlinson, B. (2010). "Who are the crowdworkers?: shifting demographics in mechanical turk." *In CHI'10 extended abstracts on Human factors in computing systems,* 2863-2872.

[28] Shearer, B. (2004). "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment." *Review of Economic Studies*, 71(2), 513-534.

Leveraging Upfront Payments to Curb Employee Misbehaviors: Evidence from a Natural Field Experiment

Online Appendix

John A. List[1], and Fatemeh Momeni[2]

A Quality

In this section, we explore whether the level of upfront payment affects the quality of transcription for the images that workers transcribe. We measure the quality of transcription using the Levenshtein distance between the text that was transcribed by a worker and the actual text on an image. The Levenshtein distance is defined as the minimum number of single character edits (i.e. insertions, deletions, or substitutions) required to convert one string of text to another. This measure can be used as a proxy for the number of errors/typos a worker makes while transcribing. Therefore, a lower Levenshtein distance corresponds to a higher quality of transcription. Table A.1 presents the average Levenshtein distance between actual text and worker's transcription across treatments.

As Table A.1 suggests, while the quality of transcription by cheaters is significantly lower than the one by non-cheaters in all treatments,[3] the overall quality does not significantly change as we vary the upfront payment across treatments.[4] We thus conclude that paying upfront does not significantly impact the quality of work.

[1]Department of Economics, University of Chicago, Chicago, IL, USA. Email: jlist@uchicago.edu.

[2]Department of Economics, University of Chicago, Chicago, IL, USA. Email: fmomeni@uchicago.edu.

[3]P-values from the Two-sample Wilcoxon rank-sum (Mann-Whitney) test against the null hypothesis of equal means for cheaters and non-cheaters: $P_{Baseline} = 0.00$; $P_{Upfront10} = 0.00$; $P_{Upfront50} = 0.00$; $P_{Upfront90} = 0.00$

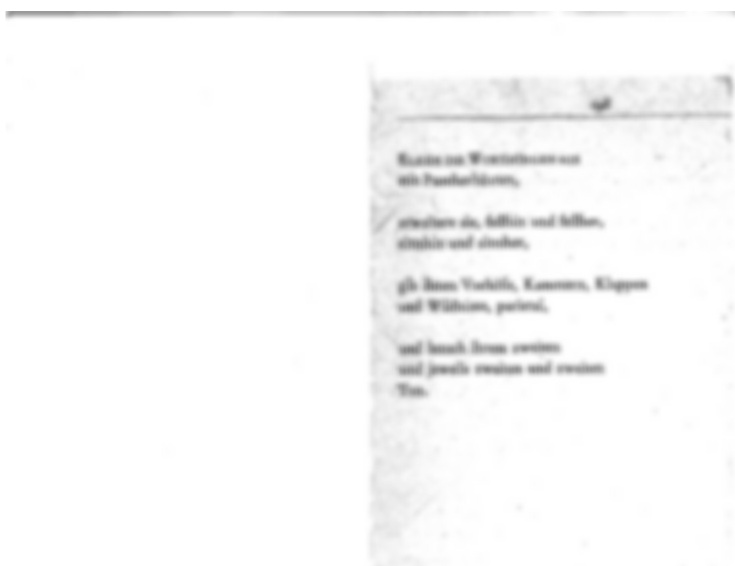
[4]P-values from the Two-sample Wilcoxon rank-sum (Mann-Whitney) test against the null hypothesis of equal means are: $P_{Baseline-Upfront10} = 0.586$; $P_{Baseline-Upfront10} = 0.684$; $P_{Baseline-Upfront10} = 0.810$; $P_{Upfront10-Upfront50} = 0.831$; $P_{Upfront10-Upfront90} = 0.732$; $P_{Upfront50-Upfront90} = 0.866$

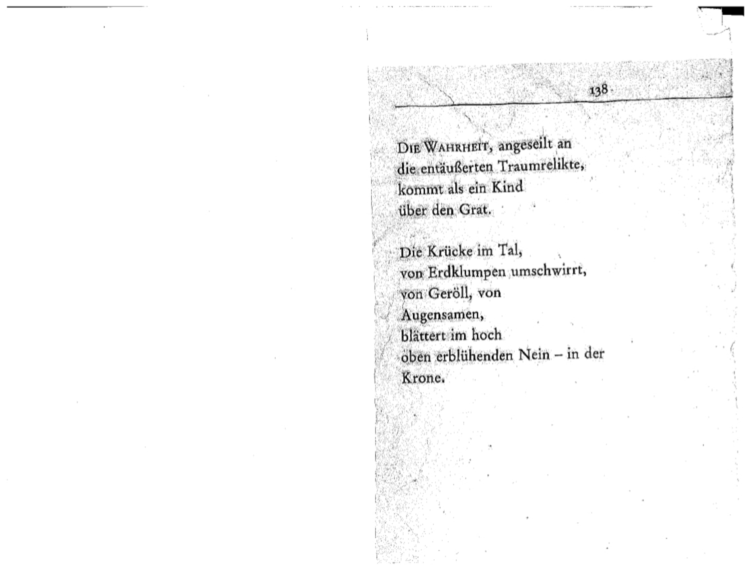
Table A.1: Quality

	Baseline	Upfront10	Upfront50	Upfront90
Mean Levenshtein Distance	23.284	22.147	22.524	23.298
(All)	(1.108)	(0.985)	(1.082)	(1.163)
Mean Levenshtein Distance	20.326	19.544	20.245	20.625
(Non-cheaters)	(0.818)	(0.680)	(0.873)	(0.937)
Mean Levenshtein Distance	35.712	42.414	36.789	35.062
(Cheaters)	(4.396)	(6.197)	(5.325)	(4.542)

Notes: Quality is defined as the minimum number of single character edits (i.e. insertions, deletions, or substitutions) required to convert one string of text to another. Standard errors are in parentheses.

B Examples of Images



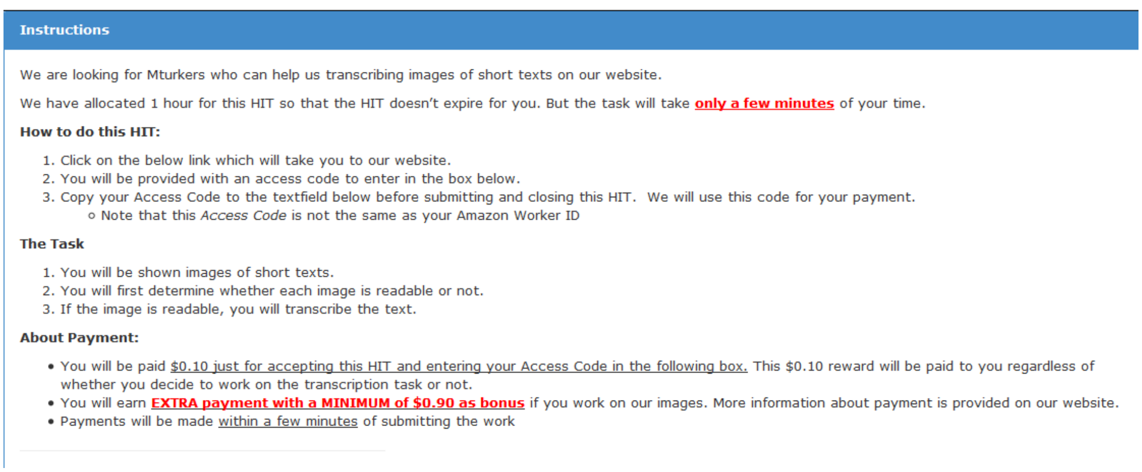


138

DIE WAHRHEIT, angeseilt an
die entäußerten Traumrelikte,
kommt als ein Kind
über den Grat.

Die Krücke im Tal,
von Erdklumpen umschwirrt,
von Geröll, von
Augensamen,
blättert im hoch
oben erblühenden Nein – in der
Krone.

C MTurk HIT



Instructions

We are looking for Mturkers who can help us transcribing images of short texts on our website.

We have allocated 1 hour for this HIT so that the HIT doesn't expire for you. But the task will take **only a few minutes** of your time.

How to do this HIT:

1. Click on the below link which will take you to our website.
2. You will be provided with an access code to enter in the box below.
3. Copy your Access Code to the textfield below before submitting and closing this HIT. We will use this code for your payment.
   - Note that this *Access Code* is not the same as your Amazon Worker ID

The Task

1. You will be shown images of short texts.
2. You will first determine whether each image is readable or not.
3. If the image is readable, you will transcribe the text.

About Payment:

- You will be paid $0.10 just for accepting this HIT and entering your Access Code in the following box. This $0.10 reward will be paid to you regardless of whether you decide to work on the transcription task or not.
- You will earn **EXTRA payment with a MINIMUM of $0.90 as bonus** if you work on our images. More information about payment is provided on our website.
- Payments will be made within a few minutes of submitting the work

D Information on the Contract

Your response to our Mturk HIT was just auto-approved and your $0.10 reward was paid on your Amazon Payment account. Please read the following information about the task and decide if you are willing to work for us for a few minutes and earn extra bonus! This HIT requires you to transcribe short texts which have been scanned from German documents.

- You will receive a total of $1.20 bonus for working on 10 images (similar to the one you see below), which will be paid as a bonus on your Amazon Payment account.

- The task is short and will not take more than a few minutes.

- If you chose to work on 10 images, we will immediately pay Y% of the total bonus to you on your Amazon Payment account upfront, and the remaining 100-Y% will be paid as soon as you complete the task (i.e. finished all 10 images).[5]

- It is possible that some of the images are too blurry to be readable. Reporting those as unreadable is acceptable and will NOT reduce your bonus payment.

- You will need to finish the task within the next 2 hours in order to receive payments.

You now have the option to choose if you want to only earn your $0.10 reward for accepting the HIT and exit; or you would like to transcribe images for us and earn an extra bonus. NOTE: Y% ($X) of the total bonus will be paid to you immediately after choosing the option to work on all 10 images and the remaining 100-Y% ($1.20-$X) will be paid within a few minutes after you complete the task. Please choose one of the following options:

- I do NOT want to transcribe images

- I want to work on all 10 images

[5] Y is treatment specific and reflects the upfront payment.