

When Corporate Social Responsibility Backfires: Theory and Evidence from a Natural Field Experiment*

John A. List[†], Fatemeh Momeni[‡]

December 2017

Abstract

Corporate Social Responsibility (CSR) has become a cornerstone of modern business practice, developing from a “why” in the 1960s to a “must” today. Early empirical evidence on both the demand and supply sides has largely confirmed CSR’s efficacy. This paper combines theory with a large-scale natural field experiment to connect CSR to an important but often neglected behavior: employee misconduct and shirking. Through employing more than 3000 workers, we find that our usage of CSR *increases* employee misbehavior – 20% more employees act detrimentally toward our firm by shirking on their primary job duty when we introduce CSR. Complementary treatments suggest that “moral licensing” is at work, in that the “doing good” nature of CSR induces workers to misbehave on another dimension that hurts the firm. In this way, our data highlight a potential dark cloud of CSR, and serve to forewarn that such business practices should not be blindly applied.

*We also thank Joseph Seidel for his 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

“Corporate social responsibility is a hard-edged business decision. Not because it is a nice thing to do or because people are forcing us to do it... but because it is good for our business” Niall Fitzgerald, Former CEO, Unilever

1 Introduction

Corporate Social Responsibility (CSR) has become one of the most common business practices in recent decades. About a quarter of N100 companies and over 90% of G250 companies now publish annual CSR reports (KPMG, 2015). While such a prevalence in CSR may seem at odds with the idea that the only responsibility of a business is to maximize its profit, recent studies have argued that CSR practices can in fact be profitable.

On the demand side, some evidence suggests that consumers value a responsibly produced good and are willing to pay a higher price for it (Du, Bhattacharya and Sen, 2011; Elfenbein et al. 2012; Lii and Lee, 2013; Sen, Bhattacharya and Korschun, 2006; Servaes and Tamayo 2013). On the supply side, workers have shown a preference for being employed by a CSR company, working harder when their employer engages in CSR (Backhaus, Stone and Heiner, 2002; Brekke and Nyborg, 2008; Greening and Turban 2000; Turban and Greening, 1997). In this spirit, companies can reduce their labor costs and increase profits through CSR. Yet, the empirical evidence on the supply-side profitability of CSR has solely focused on the effect on productivity (Burbano, 2015 and 2016; Hedblom, Hickman, & List, 2016, Tonin and Valassapoulos, 2015).

In this paper, we explore another supply-side channel through which CSR can affect profitability: the impact on employee misbehavior on the job. Employee misbehavior is a common and costly problem facing businesses and organizations. It has been estimated that companies lose about 5% of their annual revenues to various forms of internal fraud (Association of Certified Fraud Examiners, 2016). A survey from the National Retail Federation (NRF) reports that in the retail industry alone, employee theft amounted to \$15 billion (over a third of the total inventory shrinkage) in 2014 (NRF, 2015). Another survey reports that in 2015, one in every 38 employees in the retail industry was apprehended for theft from their employers.¹

Given the magnitude of the cost that organizations bear from various forms of employee misbehavior, understanding whether and how common CSR practices interact with employee misbehavior becomes of a great import. We begin with a theoretical model that highlights how CSR can influence employee misbehavior through two channels. First, it can serve as a social incentive tool for motivating workers to reduce unethical and counterproductive behavior on the job. Previous studies have shown that, consistent with the standard gift-exchange model, workers reciprocate a

¹ <http://hayesinternational.com/news/annual-retail-theft-survey/>

higher wage from their employer by reducing misbehavior on the job that hurts the employer (Flory, Liebbrandt, & List, 2016; Ockenfels, Sliwka, & Werner, 2015). Similar to how monetary-incentives reduce worker misbehavior through a gift-exchange mechanism, social incentives in the form of CSR may reduce misbehavior, by triggering reciprocity towards the employer. CSR may thus reduce worker misbehavior through the gift-exchange channel.

A second channel is that CSR can increase worker misbehavior through moral-licensing. Prosocial behavior is motivated in part, by self- (and social-) image motives: people act prosocially, in part to signal to themselves (and to others) that they are good and moral individuals (Benabou and Tirole 2006, 2011). While prosocial deeds can boost individual self-image, unethical behavior can damage it. As our moral standards are constantly challenged in multiple dimensions, good behavior in one domain may liberate us to behave unethically in another domain. Such a dynamic of moral licensing in individual behavior has been documented in the social psychology literature.² Relatedly, moral-licensing has been raised by Benabou and Tirole (2010) as a potential dark-side of CSR’s appeal to image concerns, as *“people who have recently done good in one dimension may feel immunized against negative (social or self) inferences, and thus later on act less morally constrained”*.

Our theory shows that the two channels operate in opposing directions, opening up the potential for empirical work to provide important measurement of not only the direction but the magnitude of the effect of CSR on misbehavior. To shed empirical evidence into our theory, we conduct a natural field experiment with over 3000 workers who we hired ourselves. In this manner, we served as the employer of an online labor market platform (Amazon Mechanical Turk) and invited interested workers to our website to perform a short transcription task for payment. The task and the payment structure were designed in a manner that provided opportunities for workers to misbehave. For example, all workers in our experiment received 10% of their total payment upfront, and immediately upon accepting the contract, and received the remaining 90% of their wage, once they completed the task. Receiving a percentage of the wage upfront creates an incentive for workers to accept the contract without actually working on, or completing the task.

To provide the necessary variation to identify the critical pieces of the model, we randomized workers’ into one of the 6 treatments, across which we varied wage, CSR incentive, and the framing and timing of the CSR message. Our six treatment cells provide data to answer several queries.

² A large body of experimental evidence suggests that reflecting on prior good behavior (Mazar and Zhong, 2009; Jordan, Mullen and Murnighan, 2011; Clot, Grolleau and Ibanez, 2014 and 2017; Effron, Miller and Monin, 2012; Effron, Miller and Monin 2013; Conway and Peetz, 2012; Sachdeva, Iliev and Medin, 2009) and planning to engage future good acts (Cascio and Plant, 2015; Urbszat, Herman and Polivy, 2002) can license less moral, more anti-social, and more self-indulgent choices. In addition, moral licensing has been documented in a variety of domains such as consumer behavior (Mazar and Zhong, 2009; Kouchaki and Jami, 2016), racial bias (Cascio and Plant 2015; Effron, Miller and Monin, 2012), diet and health related decisions (Effron, Miller and Monin, 2013, Urbszat, Herman and Polivy, 2002, Chiou et al., 2011), charitable giving (Conway and Peetz 2012; Sachdeva, Iliev and Medin, 2009) and energy conservation (Tiefenbeck et al. 2013). The interested reader should see Merritt, Effron and Monin, (2010) for a review.

For example, to explore the potential role of moral-licensing, we varied the appeal of CSR to workers self-image by changing the framing of our CSR messages. In addition, to understand the effect of substituting financial resources away from private incentives and into CSR, we compare behavior across treatments in which we held the total employer's budget fixed, and substituted private incentives with CSR.

Overall, our results suggest that our usage of CSR increased cheating. First, the share of workers who shirk their primary job duty increases significantly –by roughly 20%– from the baseline to our CSR treatments. Indeed, CSR not only increases the number of people who misbehave, it also increases the level of shirking: the average level of cheating by workers increases by 11%. Second, consistent with the moral-licensing effect of CSR, we find the share of cheaters to be the highest when we frame CSR as a prosocial act on behalf of workers. While CSR increases cheating, we do not find any effect on the average quality of work after accounting for the cheating. We also do not find any differences in cheating behavior across workers who could and could not sort themselves into the CSR job. We argue that our inability to document any selection effect may be due to the relatively high rate of accepting the contract by our workers in all treatments.

Combining the finding that employee misbehavior decreases with wage, documented in previous studies (Flory, Liebbrandt, & List, 2016; Ockenfels, Sliwka, & Werner, 2015), with the results from our experiment that misbehavior increases with CSR, one would expect cheating to increase as we substitute private incentives with CSR. This is in fact what our empirical analysis reveals. As we decrease the wage and increase the expenditure on CSR, we find the share of cheaters to increase considerably. In other words, we find that substituting just about 5% of the wage with CSR increases the share of cheaters by 25%, while substituting 28% of wage with CSR increases the share of cheaters by over 50%. Likewise, the intensity of cheating per worker increases as well.

In light of the fact that previous empirical studies that investigate the supply-side profitability of CSR, have generally reported positive effects on productivity (Burbano, 2015 and 2016; Hedblom, Hickman & List, 2016), our results reveal a potential dark side to the supply-side effect of CSR. We find that CSR can increase worker misbehavior on the job by generating a moral-licensing effect. Our findings also have important implications on how employers should communicate CSR initiatives to their employees. Importantly, we find some suggestive evidence that the way CSR is communicated to workers can play an important role in the extent to which it leads to moral-licensing. When communicated to the employees as a benevolent act that the employer engages in, on workers' behalf, CSR is (marginally insignificantly) more likely to exploit workers self-image and to lead to less moral choices in the future. This result is consistent with the findings of Kouchaki and Jami (2016) who document a higher level of moral-licensing, when consumers are exposed to a CSR message that praises the consumers compared to a message that praised the company for CSR.

Our study makes important contributions to the recently growing literature on moral-licensing. While most previous evidence on moral-licensing is limited to evidence from lab, our study is one of the few that finds moral-licensing in a natural field setting, where subjects are not aware that they are taking part in the study.³ The scale of our experiment is another factor that distinguished our study from previously documented evidence on moral-licensing. In a meta-analysis of 91 studies in moral licensing, Blanken, van de Ven and Zeelenberg (2015) finds the sample size in these studies to be relatively small with the average sample size of only 81.3 participants per study. With a sample size of over 3000 subjects, our experiment is one of the largest scale in which moral-licensing has been documented so far.

The remainder of our study is organized as follows. Section 2 presents our theoretical framework. Section 3 lays out our experimental design, and Section 4 presents our findings. Section 5 concludes.

2 Theoretical Framework

In this section, we present a simple framework that enables us to formalize the two channels through which CSR can affect misbehavior on the job. Consider a worker in a firm that pays a wage of W and invests in CSR. For the sake of simplicity, we assume that the worker makes only one decision: the level of cheating X , which can reflect various measures of misbehavior on the job that are harmful to the employer, such as stealing from, or lying to the employer.^{4,5} Thus, employer’s payoff is decreasing in X . The worker observes W and CSR, set by the employer, and chooses X by maximizing the following utility function:

$$U(X, W, CSR) = W - G(X, CSR) + B(X) + A(W, CSR) \times (-X)$$

$B(X)$ represents the benefit to the worker from cheating at level X .⁶ We assume that the worker cares about his self-image. Unethical behavior (X) damages worker’s self-image, and CSR, especially if perceived as a prosocial act by the worker, improves it. Alternatively, we could think of self-image as an inverse function of guilt $G(X, CSR)$. The worker’s feeling of guilt increases with his unethical choice (X) and decreases with CSR. The worker also exhibits gift-exchange motives towards the employer and cheats less if he perceives the employer to be of a “kind” type. We

³ Tiefenbeck et al. (2013) and study 2 in Kouchaki and Jami (2016) are exceptions.

⁴In this simplified model, worker’s effort can be thought of as an inverse function of cheating, in a sense that more effort corresponds to less cheating. We interchange shirking, cheating, and misbehavior throughout the study.

⁵We use the following definition of unethical misbehavior taken from Kish-Gephart et al., (2010). “*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*”.

⁶ Alternatively, $-B(X)$ can be thought of as the cost of reducing cheating, or exerting effort on the job.

assume that both wage and CSR can increase the worker’s perception of the employer’s level of kindness $A(W, CSR)$.⁷ Thus, an employer is perceived to be kinder if it pays a higher wage, or spends more on CSR.

The first and the second order conditions to the workers problem can be expressed as:

$$\text{FOC: } -\frac{\partial G(X, CSR)}{\partial X} + \frac{\partial B(X)}{\partial X} = A(W, CSR)$$

$$\text{SOC: } -\frac{\partial^2(X, CSR)}{\partial X^2} + \frac{\partial^2 B(X)}{\partial X^2} < 0$$

We can use the Implicit Function Theorem to show: $\frac{dX}{dCSR} = \frac{\frac{\partial}{\partial CSR}(\frac{\partial G}{\partial X}) + \frac{\partial A}{\partial CSR}}{-\frac{\partial^2 G}{\partial X^2} + \frac{\partial^2 B}{\partial X^2}}$. Following the second order condition, the denominator of this expression is negative. Thus, we can drive the following conditions for how misbehavior responds to the change in CSR:

- $\frac{dX}{dCSR} > 0 \rightarrow \frac{\partial}{\partial CSR}(\frac{\partial G}{\partial X}) + \frac{\partial A}{\partial CSR} < 0$
- $\frac{dX}{dCSR} < 0 \rightarrow \frac{\partial}{\partial CSR}(\frac{\partial G}{\partial X}) + \frac{\partial A}{\partial CSR} > 0$
- $\frac{dX}{dCSR} = 0 \rightarrow \frac{\partial}{\partial CSR}(\frac{\partial G}{\partial X}) + \frac{\partial A}{\partial CSR} = 0$

The above conditions suggest that depending on the specific functional form of the gift-exchange and the guilt term, as CSR increases, worker misbehavior may increase or decrease. Specifically, (since by assumption, A is an increasing function of CSR), misbehavior increases with CSR if the cross partial derivative of $G_{X, CSR}$ is negative and is larger in magnitude than the slope of A with respect to CSR, $(\frac{\partial A}{\partial CSR})$. In other words, misbehavior increases with CSR if, as CSR increases, the change in the slope of the guilt function (G) with respect to misbehavior (X) decreases with a rate larger than the rate with which A increases.

3 Experimental Design

We conduct our study as a natural field experiment (see Harrison and List, 2004), using workers from Amazon’s Mechanical Turk (MTurk), an online labor market platform where businesses and individuals can post tasks and workers can work on tasks to receive payment. We acted as an employer and recruited workers on MTurk for a short task to be performed on our website. The task was designed such that it provided opportunities for workers to cheat on, without losing their payment, or risking their reputation on MTurk. We invited interested workers to our website where we randomized them into one of our treatments, and offered them a short “contract” with treatment-specific information. To get at how CSR affects cheating on the job, we randomized

⁷ Previous research has shown that financial and non-financial incentives can both induce gift-exchange in workers (Bradler, et al., 2016; Dur, 2009; Kosfeld & Neckermann, 2011; Kube, Maréchal, & Puppe, 2012).

workers into treatments across which we fixed the wage and varied the CSR incentive. To explore how substituting private incentive with CSR affects cheating, we fixed the total employer’s budget across 3 treatments and varied the level of wage and CSR expenditure.

Recruitment on MTurk. The MTurk environment provides an appropriate platform to study misbehavior on the job. The relative anonymity and the remote nature of the job makes it possible to design a task with possibilities of cheating. We recruited workers on MTurk by posting a Human Intelligence Task (HIT) and invited interested workers to our website where they could learn more about the task and payment, and worked on the task. The HIT (Figure 1) provided a general description of the task and instructions for the interested workers on how to proceed. Interested workers were instructed to click on a link that took them to our page where a unique ID-code was generated for each worker. Workers were then required to submit their code through MTurk. We later used these ID-codes to link workers on our website to their MTurk account, through which we made payments.⁸

The Contract. Upon landing on our website, workers were randomized into one of the 6 treatment groups and were provided with a “contract” with detailed information about the task and treatment-specific payment. The contracts indicated that 10% of the total wage would be paid to workers upfront, upon accepting the contract, and the remaining 90% would be paid after they completed the task. After reading the information, workers had to decide whether or not to accept the contract. Workers were explicitly asked to accept the contract, only if they intended to complete the task. All workers who accepted the contract were immediately paid the upfront money and were sent an email through MTurk, notifying them about the upfront payment. Accepting the contract (and consequently receiving the upfront payment) without finishing the task serves as one of our measures of worker misbehavior. The contract required workers to complete the task within 2 hours of accepting the contract.

Treatment Groups. Workers were randomized into one of the 6 treatment groups illustrated in Table 1.

Table 1: Treatments

	$T_{Baseline}^{MWage}$ (1)	$T_{Baseline}^{HWage}$ (2)	$T_{FirmMsg}^{MWage}$ (3)	$T_{FirmMsg}^{LWage}$ (4)	$T_{WorkerMsg}^{MWage}$ (5)	T_{Sort}^{MWage} (6)
Wage	\$1.20	\$1.26	\$1.20	\$0.90	\$1.20	\$1.20
CSR as % of wage bill	0	0	5%	40%	5%	5%
CSR msg	-	-	Firm	Firm	Worker	Firm
CSR+Wage	\$1.20	\$1.26	\$1.26	\$1.26	\$1.26	\$1.26
Sorting	-	-	No	No	No	Yes

⁸ The codes were also used to ensure that each worker was only exposed to one treatment.

Depending on the treatment, workers were offered either a low wage of \$0.90, a medium wage of \$1.20, or a high wage of \$1.26 for completing the task.⁹ In addition to wage, workers in treatments 3 to 6, also received a social incentive in the form of CSR. Our CSR initiative was in the form of donating cash to a pre-specified non-profit organization. Donation to charities and non-profit organizations is a form of CSR that is frequently advertised and practiced by firms and companies.¹⁰ Companies often report their charity donations as the equivalent of a percentage of their profit. However, as we did not make any profits from our experiment, in order to avoid deception, we reported our donation as the equivalent of a percentage of our total wage-bill.

As previously discussed in section 2, if the moral-licensing effect of CSR is strong enough, CSR can increase cheating. Moral-licensing works through concerns for self-image. If workers see themselves as a part of the prosocial act that their employer engages in, CSR can improve workers' self-image and license less ethical subsequent acts. The ability to take credit for the prosocial act of the employer is an important determinant of whether CSR influences the worker's self-image and therefore is key in whether it leads to moral-licensing. The way in which CSR is framed and communicated to employees can influence the extent to which workers' self-image is exploited by CSR. To shed light on the role of moral-licensing in the relationship between CSR and misbehavior, we manipulated the appeal of CSR to workers' self-image by changing the framing of our CSR message across treatments. While for some workers ($T_{FirmMsg}^{LWage}$, $T_{FirmMsg}^{MWage}$ and T_{Sort}^{MWage}) we simply framed CSR as "donation to the charity", for others ($T_{WorkerMsg}^{MWage}$) we framed it as "donation to the charity on behalf of the workers". If CSR is to increase cheating through moral-licensing, we should expect the later framing to generate a stronger effect. The exact wordings of our CSR message read as following:

*"Our firm is committed to give back in meaningful ways. We are passionate about encouraging education for the next generation. We do our part by donating money to influential non-profit organizations that support education for children from low socioeconomic backgrounds. In keeping with our philanthropic mission, we donate the equivalent of x% of our wage bill in cash (on behalf of all workers who help us with this project) to UNICEF Education Programs. UNICEF works tirelessly to ensure that every child regardless of gender, ethnicity or circumstances has access to a quality education. You may find out more about UNICEF Education Programs at: UNICEF."*¹¹

The information about CSR was given to workers at different stages of the experiment, depending on the treatment. While in the CSR treatments without the possibility of sorting ($T_{FirmMsg}^{MWage}$,

⁹The average hourly earnings on MTurk has been estimated to lie between \$1 and \$5. Given that an average worker spent about 20 minutes on our website, the average hourly earnings of our workers falls within this estimated range (Ross et al., 2010; Paolacci et al., 2010; Berg, 2016).

¹⁰To name a few examples, in 2015 alone, Walmart, Wells Fargo, Chevron, Target and Microsoft made about \$301 million, \$281 million, \$225 million, \$111 million and \$135 million of cash donations to nonprofits (see: <http://fortune.com/2016/06/22/fortune-500-most-charitable-companies/>)

¹¹We provided a link to UNICEF's web-page. The text in the parenthesis was added only in $T_{WorkerMsg}^{MWage}$.

$T_{FirmMsg}^{LWage}$ and $T_{WorkerMsg}^{MWage}$), workers were provided with the CSR information after they accepted the contract and before they started to work on the task, in our sorting treatment, T_{Sort}^{MWage} , subjects saw the CSR message on the first page on our website, and before they decided whether to accept the contract. Thus, workers in the sorting treatment T_{Sort}^{MWage} could select themselves into working for a CSR-type employer, while the other three CSR treatments removed the possibility of sorting. Therefore, the comparison of workers' behavior across $T_{FirmMsg}^{MWage}$ and T_{Sort}^{MWage} enables us to distinguish the selection effect of CSR on worker misbehavior from its direct treatment effect.

Finally, to shed light on how substituting wage with CSR affects cheating, we fixed the total employer's budget and varied the level of wage and CSR expenditure across $T_{baseline}^{HWage}$, $T_{FirmMsg}^{MWage}$ and $T_{FirmMsg}^{LWage}$. While in $T_{baseline}^{HWage}$ the entire budget was spent on incentivizing workers through wage, in $T_{FirmMsg}^{MWage}$ and $T_{FirmMsg}^{LWage}$, around 95% and 71% of the total budget was allocated to wage, and the remaining of it was spent on CSR. By comparing cheatings as we move away from $T_{baseline}^{HWage}$ to $T_{FirmMsg}^{MWage}$ and $T_{FirmMsg}^{LWage}$, we are going to explore how allocating resources away from private financial incentives and into CSR affects misbehavior on the job.

The Task. We provided each worker with a batch of 10 images of short German texts, scanned from old German books to transcribe.¹² On average, each image was composed of around 30 words or 183 characters.¹³ We used German texts to make the task harder and less enjoyable for our predominantly non-German workers.¹⁴ Workers who submitted all 10 images received a full wage of W , which was specified in the contract.

Before starting to work on any given image, workers were required to report if the image was legible or not.¹⁵ If an image was reported as unreadable, the worker skipped that image and moved on to the next. The instructions made it clear that it was possible for some images to be illegible, and that reporting those as unreadable was acceptable and did not lower the payment. Offering fixed payments, regardless of the number of images that were actually transcribed, provided opportunities for workers to misreport perfectly readable images as unreadable and avoid transcribing them. We consider such behavior as another measure of misbehavior on the job. We thus use two measures of cheating to analyze and compare across treatments. Type-1 cheating refers to misreporting readable images as unreadable, a form of shirking. Type-2 cheating refers to accepting the contract (and receiving the upfront payment) without completing the job—essentially “taking the money and running”.

¹²Transcription is among the most common tasks that are posted by requesters on MTurk.

¹³To make it harder for workers to use transcription softwares which automatically transcribe images of texts, before scanning each text that was printed on paper, we crinkled the papers to slightly deform the shape of each character on the text.

¹⁴We restricted our HIT to workers from the US and India. American and Indians make up the largest share of the workers on MTurk.

¹⁵ A small share (10%) of images were made too blurry to be readable, and the remaining 90% were made to be easily readable.

Payment. Employers on MTurk have two means of making payments to workers. Each employer is required to set a “reward” for each assignment that they post on MTurk. A reward is a pre-specified level of payment by the employer (also known as requester) that will be paid to workers if (and as soon as) the employer approves the submitted HIT. In addition to the reward, an employer can make extra payments as “bonus”. We set the reward for our assignments to \$0.10, which was paid to all workers who submitted the HIT, before they were randomized into treatments and were provided with the contract.¹⁶ The treatment-specific wage that was specified in contracts was paid to the workers who accepted the contracts through “bonuses”. Using *bonuses* as our method of paying wages, we were able to offer contracts with different wages to our workers in different treatments.

Reputational Concerns. A requester on MTurk may limit his/her HITs to workers who meet certain requirements. Once a worker completes a task for the requester, the requester may review the work and either approve or reject the submission. Future requesters may limit their tasks to workers whose lifetime approval rate is above a certain desired level. The approval rate may thus serve as a signal of quality of a worker and can be used by requesters as a requirement that potential workers have to meet. In order to eliminate potential reputational concerns and increase the baseline cheating rate, we automatically approved all the HITs that were submitted by workers on MTurk, before offering the contract. We notified workers that their submission was approved through automated messages that we sent to them through MTurk.

In summary, the timeline of the experiment was as follows: 1) recruitment on MTurk (link to our webpage); 2) contract offered on the first page of our website (CSR message if treatment = T_{Sort}^{MWage}); 3) decision to accept or reject the contract; 4) payment of 10% of the wage, if contract was accepted; 4) task instructions (CSR message if treatment = $T_{FirmMsg}^{MWage}$, $T_{FirmMsg}^{MWage}$ or $T_{WorkerMsg}^{MWage}$); 5) transcription task; 6) payment of the remaining of the wage.

4 Experimental Results

The experiment was conducted between late October 2016 and early January 2017. We collected data until we had at least 500 subjects who accepted the contract in each treatment cell.¹⁷ In total, 3188 workers landed on our website and were randomized into one of our 6 treatments. Out of these workers, 3022 accepted the contracts they were offered, with the average acceptance rate of 94.79% across treatments. On average, each worker spent about 20 minutes on our website.

¹⁶We automatically approved the HITs that were submitted by all workers.

¹⁷ We observed an average of about 22% cheating rate in our pilot. Our power calculation suggested that we would need around 450 subjects in each treatment in order to be able to detect about 30% change in mean cheating from the baseline (power=0.80; alpha=0.05) (List, Sadoff & Wagner, 2011).

In this section, we first present and discuss our findings with regard to how CSR affects worker misbehavior. Then, we explore how workers respond to the substitution of private incentives with CSR.

4.1 How Does CSR Affect Employee Misbehavior?

To shed light on how CSR affects cheating on the job, we compare workers' choices across treatments $T_{baseline}^{MWage}$, $T_{FirmMsg}^{MWage}$, $T_{WorkerMsg}^{MWage}$ and T_{Sort}^{MWage} . We start by comparing the propensity to accept the contract offer across these treatments. Note that since up to the point of making the decision of whether to accept the contract, workers in treatments $T_{baseline}^{MWage}$, $T_{FirmMsg}^{MWage}$ and $T_{WorkerMsg}^{MWage}$ have been exposed to the exact same information and have been offered the same contract, we do not expect any differences in acceptance rates across these three treatments. Workers in T_{Sort}^{MWage} however, are exposed to the CSR message, prior to making the decision to accept or reject the contract. Thus, T_{Sort}^{MWage} workers may select themselves into the job, based on their preferences for CSR. As reported on Table 2, the acceptance rates in these four treatments lie between 94.59% to 96.01%. The pairwise comparisons of acceptance rates reveals no differences across treatments $T_{baseline}^{MWage}$, $T_{FirmMsg}^{MWage}$, $T_{WorkerMsg}^{MWage}$ and T_{Sort}^{MWage} (p-values from two-sample test of proportions between pairs of treatments: $P_{1-3} = 0.71$; $P_{1-5} = 0.45$; $P_{1-6} = 0.88$; $P_{2-4} = 0.27$; $P_{3-5} = 0.27$; $P_{3-6} = 0.83$; $P_{5-6} = 0.38$). The insignificance of the difference in acceptance rates across treatments with and without the possibility of sorting suggests that our ability to document any selection effects may be limited.

Table 2: Acceptance Rates in Treatments 1, 3, 5 and 6

	$T_{Baseline}^{MWage}$ (1)	$T_{FirmMsg}^{MWage}$ (3)	$T_{WorkerMsg}^{MWage}$ (5)	T_{Sort}^{MWage} (6)
Number of workers who landed on our website and were randomized into a treatment	528	536	526	527
Number of workers who accepted the contract	502	507	505	500
Acceptance Rate	95.08%	94.59%	96.01%	94.88%

Next, we compare workers' cheating behavior across $T_{baseline}^{MWage}$, $T_{FirmMsg}^{MWage}$, $T_{WorkerMsg}^{MWage}$ and T_{Sort}^{MWage} . We study cheating both in the extensive margin, and in the intensity of cheating by each worker.

4.1.1 Share of Cheaters

Does CSR affect the share of workers who cheat? To answer this question, we compare the share of cheaters in the baseline to the one in our CSR treatments. We refer to a worker as a *cheater* if he or

she commits either type-1 or type-2 cheating, that is, he or she either misreports a readable image as unreadable, or accepts the contract and receive the upfront payment without completing the job. We find CSR to increase the share of cheaters by about 24%. The share of workers who cheat significantly increases from 0.209 (std. err. = 0.018) in the baseline $T_{baseline}^{MWage}$ to 0.259 (std. err. = 0.014) in the two CSR treatments without the possibility of sorting: $T_{FirmMsg}^{MWage}$ and $T_{WorkerMsg}^{MWage}$ (p-value =0.034).¹⁸

Consistent with the moral-licensing effect (as shown in Figure 2 and Table 3), framing CSR as a benevolent act *on behalf of the workers* ($T_{WorkerMsg}^{MWage}$) leads to a larger increase (around 30%) in the share of cheaters, than simply framing it as employer’s donation ($T_{FirmMsg}^{MWage}$). The share of cheaters increases by 18% from 0.209 (std. err. = 0.018) in the baseline $T_{baseline}^{MWage}$ to 0.247 (std. err. = 0.019) in the CSR firm framing treatment $T_{FirmMsg}^{MWage}$ (p-value=0.157), while it increases by about 30% from baseline to 0.271 (std. err. = 0.020) in $T_{WorkerMsg}^{MWage}$ (p-value=0.021). However, the difference in the share of cheaters across the two farming of CSR is not large enough to be significant given our sample size. When workers can sort themselves into working for a CSR employer, the share of cheaters is not significantly different from that of the baseline ($T_{baseline}^{MWage}$) or the corresponding treatment without the possibility of sorting ($T_{FirmMsg}^{MWage}$). Our inability to document a selection effect is consistent with our previous finding of indistinguishable acceptance rates across the treatment with and without the possibility of sorting.

Table 3: Effect of CSR on the share of cheaters

	$T_{Baseline}^{MWage}$ (1)	$T_{FirmMsg}^{MWage}$ (3)	$T_{WorkerMsg}^{MWage}$ (5)	T_{Sort}^{MWage} (6)
Share of cheaters among those who accepted the contract	0.209 (0.018)	0.247 (0.019)	0.271 (0.020)	0.226 (0.019)
Share of type-1 cheaters among those who accepted the contract	0.056 (0.010)	0.073 (0.012)	0.101 (0.013)	0.076 (0.019)
Share of type-2 cheaters among those who accepted the contract	0.159 (0.016)	0.181 (0.017)	0.174 (0.016)	0.158 (0.016)
Share of cheaters among those who accepted the contract and finished the task	0.059 (0.012)	0.080 (0.013)	0.118 (0.016)	0.081 (0.013)

Next, we explore the effect of CSR on the share of type-1 and type-2 cheaters, separately. The share of type-1 cheaters significantly increases by over 55% from 0.056 (std. err. = 0.010) in the baseline to 0.087 (std. err. =0.009) in the two CSR treatments with no sorting, $T_{FirmMsg}^{MWage}$ and $T_{WorkerMsg}^{MWage}$ (p-value =0.032). Looking at the effect of each CSR framing separately, we find CSR to increase the share of type-1 cheaters by over 30% to 0.073 (str. err. 0.012) in $T_{FirmMsg}^{MWage}$, and

¹⁸ The p-values reported in all extensive margin analysis are from Two-sample Test of Proportions against the Null hypothesis of equal means.

further to 0.101 (str. err. 0.013) in $T_{WorkerMsg}^{MWage}$. While the increase from $T_{baseline}^{MWage}$ to $T_{FirmMsg}^{MWage}$ is insignificant, (p-value=0.266), the increase from the baseline to $T_{WorkerMsg}^{MWage}$ is large and significant (p-value=0.008). Framing CSR as prosocial act on behalf of workers nearly doubles the number of type-1 cheaters.¹⁹ Selection does not seem to play a major role, when we look at the share of type-1 cheaters separately, either. The share of type-1 cheaters in T_{Sort}^{MWage} is 0.076 (std. err. = 0.019) which is not significantly different from the one in $T_{FirmMsg}^{MWage}$, where sorting is not possible. While the share of type-2 cheaters also increases with CSR from 0.159 (std. err. = 0.016) in $T_{baseline}^{MWage}$ to 0.178 (std. err. = 0.012) in $T_{FirmMsg}^{MWage}$ and $T_{WorkerMsg}^{MWage}$, the increase is not significant (p-value=0.367).

The above finding that cheating in the extensive margin increases with CSR is robust to restricting our analysis to workers who completed the task and thus received full payments. The share of cheaters among those who finished the task and submitted all 10 images significantly increases from 0.059 (std. err. = 0.012) in $T_{baseline}^{MWage}$ to 0.099 (std. err. = 0.010) in $T_{FirmMsg}^{MWage}$ and $T_{WorkerMsg}^{MWage}$ (p-value=0.019).²⁰ Again, the increase in the share of cheaters is the largest, when CSR is framed as a benevolent act on behalf of the worker (Figure 3).²¹

Result-1: CSR increases the share of cheaters. Consistent with moral-licensing, the share of cheaters is the highest when CSR is framed as a prosocial act on behalf of the worker.

4.1.2 Intensity of Cheating

Our extensive margin analysis suggested that CSR turns more workers into cheaters. We now turn into how CSR affects the intensity of cheating. Does CSR increase the overall level of cheating as well? To answer this question, we compare the level of cheating by workers across the baseline and CSR treatments, measured by the number of images that a worker either misreported as unreadable (type-1 cheating), or did not submit (type-2 cheating).

We find CSR to increase the average intensity of cheating per worker by about 11%, from 1.540 (std. err. = 0.150) images in the baseline to 1.716 (std. err.=0.107) in $T_{FirmMsg}^{MWage}$ and $T_{WorkerMsg}^{MWage}$ (p-value =0.06).²² Exploring the effect of each type of CSR messaging separately, the intensity of cheating increases from the baseline by about 13% to 1.740 in $T_{FirmMsg}^{MWage}$ (p-value=0.171) treatment and by about 10% to 1.693 in $T_{WorkerMsg}^{MWage}$ (p-value=0.061). Note that while the increase in cheating in $T_{FirmMsg}^{MWage}$ is larger in magnitude, the increase is only significant in $T_{WorkerMsg}^{MWage}$. Similar to the

¹⁹ The difference in the share of type-1 cheaters between the two framings of CSR message is marginally insignificant (p-value=0.11).

²⁰ Note that by definition, those who finished the task can only commit type-1 cheating.

²¹ The share of cheaters who finished the task when selection was possible is 0.081 (std. err. = 0.013), not significantly different from the treatment without sorting.

²² 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.

extensive margin results, selection into CSR job does not seem to play a role in the intensity of cheating: the average cheating per worker is 1.488 (std. err. = 0.145) in T_{Sort}^{MWage} which is not significantly different from the baseline, $T_{baseline}^{MWage}$ (p-value=0.684) or from the corresponding treatment with no sorting $T_{FirmMsg}^{MWage}$ (p-value=0.328). Figure 4 and Table 4 summarize these findings.

Table 4: Effect of CSR on the intensity of cheating

	$T_{Baseline}^{MWage}$	$T_{FirmMsg}^{MWage}$	$T_{WorkerMsg}^{MWage}$	T_{Sort}^{MWage}
	(1)	(3)	(5)	(6)
Total cheating: average number of images a worker cheated on	1.540 (0.150)	1.740 (0.156)	1.693 (0.148)	1.488 (0.145)
Type-1 cheating: average number of readable images a worker misreported as unreadable	0.145 (0.038)	0.162 (0.034)	0.251 (0.045)	0.150 (0.032)
Type-2 cheating: average number of images NOT submitted per worker	1.394 (0.147)	1.578 (0.154)	1.441 (0.145)	1.338 (0.143)
Average number of readable images a worker misreported as unreadable among workers who completed the task	0.159 (0.044)	0.181 (0.041)	0.290 (0.053)	0.164 (0.037)

Focusing on the effect of CSR on the intensity of type-1 cheating, we find a 43% increase from 0.145 (std. err. = 0.038) images per worker in the baseline to 0.207 (std. err. = 0.028) images per worker in the CSR treatments $T_{FirmMsg}^{MWage}$ and $T_{WorkerMsg}^{MWage}$ (p-value=0.031). Examining the effect of each type of CSR messaging separately, we find the intensity of type-1 cheating to increase from 0.145 in the baseline to 0.162 (std. err. =0.034) in $T_{FirmMsg}^{MWage}$ (p-value=0.267) and to 0.252 (std. err. = 0.045) in $T_{WorkerMsg}^{MWage}$ (p-value=0.007). Once again, selection does not play a role: the intensity of type-1 cheating by workers who can sort themselves into the CSR job is 0.150 (std. err. = 0.032) images per worker, not significantly different from the baseline (p-value=0.207), or $T_{FirmMsg}^{MWage}$ (p-value=0.89). While the intensity of type-2 cheating also increases with CSR, from 1.394 in the baseline to 1.578 (std. err. = 0.154) and 1.441 (std. err. = 0.145) in treatments $T_{FirmMsg}^{MWage}$ and $T_{WorkerMsg}^{MWage}$, the increase is not significant. Therefore, the increase in the intensity of cheating is mainly driven by type-1 cheating.

The conclusions are robust to restricting the sample to workers who finished the task and received full payment: as graphed in Figure 5, CSR significantly increases the intensity of overall cheating by about 48% from 0.159 in the baseline to 0.236 (std. err. = 0.033) in $T_{FirmMsg}^{MWage}$ and $T_{WorkerMsg}^{MWage}$ (p-value=0.018). Decomposing the effect of each type of CSR framing, we find a 14% increase in cheating from the baseline to 0.181 (std. err. = 0.041) in $T_{FirmMsg}^{MWage}$ (p-value = 0.249) and an 82% increase in cheating from the baseline to 0.290 (std. err. = 0.053) in $T_{WorkerMsg}^{MWage}$ (p-value = 0.003).

Result-2 summarizes our findings on the effect of CSR on the intensity of cheating by workers.

Result-2: CSR increases the intensity of cheating. In-line with the moral-licensing effect of CSR, workers lie more about legibility of an image when CSR is framed as a prosocial act on behalf of the worker.

4.1.3 Intensity of Cheating by Cheaters

Our results presented in section 4.1.1, suggest that CSR can turn some non-cheaters into cheaters. Here we delve deeper into understanding the behavior of cheaters across treatments. Specifically, we explore whether a worker who cheats under CSR treatment behaves differently from a worker who cheats in the baseline. To answer this question, we compare the intensity of cheating by cheaters in the baseline and CSR treatments. While the previous section demonstrated that the intensity of cheating increases with CSR, when we focus on the behavior of cheaters, we find the intensity of cheating by cheaters to decrease with CSR (Table 5 and Figure 6). The average cheating per cheater falls from 7.362 (std. err. = 0.326) in the baseline to 6.630 (std. err. = 0.217) in the two CSR treatments with no sorting, $T_{FirmMsg}^{MWage}$ and $T_{WorkerMsg}^{MWage}$ (p-value=0.094).

Table 5: The intensity of cheating by cheaters

	$T_{Baseline}^{MWage}$ (1)	$T_{FirmMsg}^{MWage}$ (3)	$T_{WorkerMsg}^{MWage}$ (5)	T_{Sort}^{MWage} (6)
Total cheating: average number of images a cheater cheated on	7.362 (0.326)	7.056 (0.314)	6.241 (0.298)	6.584 (0.335)
Type-1 cheating: average number of readable images a type-1 cheater misreported as unreadable	2.607 (0.494)	2.216 (0.310)	2.490 (0.297)	1.974 (0.291)
Type-2 cheating: average number of images NOT submitted by each type-2 cheater	8.750 (0.215)	8.696 (0.213)	8.273 (0.219)	8.468 (0.229)
Average number of readable images a cheater misreported as unreadable among cheaters who submitted all 10 images	2.680 (0.550)	2.273 (0.346)	2.469 (0.303)	2.029 (0.320)

Examining the effect of each type of CSR message separately, we find the average cheating by cheaters to fall from the baseline to 7.056 (std. err. = 0.314) in $T_{FirmMsg}^{MWage}$ and further to 6.241 (std. err. = 0.298) in $T_{WorkerMsg}^{MWage}$. These findings suggest that the average cheater in the CSR treatments cheats less than the average cheater in the baseline. Figure 8 graphs the fraction of cheaters in the baseline and the CSR treatments ($T_{FirmMsg}^{MWage}$ and $T_{WorkerMsg}^{MWage}$) by different levels of cheating. If we call cheaters who cheat on fewer than 4 images, low-type cheaters and those who cheat on more than 7 images, high-type cheaters, we can see that CSR increases the fraction of low-type cheaters,

and decreases the fraction of high-type cheaters. This pattern becomes even clearer, if we compare the fraction of low-type and high-type cheaters in the baseline to the one in $T_{WorkerMsg}^{MWage}$ (Figure 8). Note that despite the fall in cheating by the cheaters, as we showed in the previous section, the overall cheating still increases with CSR, suggesting that the extensive margin effect dominates. We summarize these findings in Result-3.

Result-3: While CSR induces more workers to cheat, an average worker who cheats under CSR, cheats less often than an average worker who cheats in the baseline.

4.1.4 Quality

So far, we have shown that CSR increases cheating: workers who are exposed to the CSR message are more likely to either lie about the legibility of images or to receive the upfront payment but not complete the task. In this section, we explore whether CSR affects the quality of work for those images that workers do transcribe. We measure the quality of transcription using the Levenshtein distance between the texts that was transcribed by the workers and the actual text on the images. The Levenshtein distance is defined as the minimum number of single character edits (i.e. insertions, deletions or substitutions) required to go from one string of text to another. We can think of this measure as the number of errors/typos that a worker makes while transcribing. Therefore, a lower Levenshtein distance corresponds to a higher quality of transcription.

Table 6: Quality of transcription

	$T_{Baseline}^{MWage}$ (1)	$T_{FirmMsg}^{MWage}$ (3)	$T_{WorkerMsg}^{MWage}$ (5)	T_{Sort}^{MWage} (6)
Mean Levenstein distance	22.147 (0.985)	22.795 (0.984)	23.641 (1.124)	22.950 (1.001)
Mean Levenstein distance for non-cheaters	20.864 (0.940)	21.585 (0.893)	21.957 (0.978)	21.694 (0.857)
Mean Levenstein distance for cheaters	42.414 (6.197)	34.853 (3.658)	37.455 (4.253)	30.407 (4.107)

As Table 6 illustrates, the average Levenstein score in baseline is 22.147 and increases to 22.795 and 23.641 in $T_{FirmMsg}^{MWage}$ and $T_{WorkerMsg}^{MWage}$. None of these increases are significant. Looking at the quality by cheaters versus non-cheaters, one can see that cheaters' work is of a much lower quality, across all treatments.²³ The number of errors made by cheaters is on average between 43% to 110% higher than the ones made by non-cheaters across treatments. Sorting does not seem to play a role either.

Result-4: There is no significant difference in the quality of transcription across treatments. Cheaters'

²³ The difference in work quality by cheaters and non-cheaters is significant in 0.1% level in all treatments.

work is of a much lower quality across all treatments.

To summarize, we find CSR to increase cheating. Importantly, we find that CSR increases cheating the most when it has a greater appeal to workers' self-image concern. This observation is consistent with moral-licensing that suggests a boost in self-image through a benevolent act can license less ethical choices in the future. While we find CSR to increase misbehavior, we do not find any effect on the quality of work delivered by workers. Quality of work often serves as an important dimension of a worker's behavior. While low quality work can be counterproductive and costly to the employer, a worker who delivers low quality work may not experience as much guilt as a worker who engages in other forms of counterproductive behavior such as stealing or lying. Since we expect moral-licensing to liberate the behavior that are otherwise considered unethical and lower self-image, CSR may not lead to moral-licensing in the quality dimension as strongly, if low-quality work does not affect worker's self-image.

4.2 Pay for Wage or Pay for CSR?

In this section, we investigate how worker's behavior changes, as employer substitutes the private incentive (wage) with CSR. Results from previous studies suggest that, consistent with the prediction of the gift-exchange model, increasing wage reduces misbehavior on the job (Flory et al., 2016; Ockenfels et al., 2015). Putting these results together with our finding from the previous section that CSR increases cheating, we expect misbehavior to increase, as we allocate resources away from wage and into CSR. This is indeed what our empirical analysis reveals. Below, we report our findings, in terms of both the share of cheaters and the intensity of cheating.

We investigate how workers' behavior change as wage is substituted with CSR, by comparing choices across $T_{baseline}^{HWage}$, $T_{FirmMsg}^{MWage}$ and $T_{FirmMsg}^{LWage}$. Across these three treatments, we keep our total employer's expenditure on wage and CSR fixed, and reduce the share of the budget spent on wage as we move from $T_{baseline}^{HWage}$ to $T_{FirmMsg}^{MWage}$ and $T_{FirmMsg}^{LWage}$. While in $T_{baseline}^{HWage}$, the entire budget is spent on wage, in $T_{FirmMsg}^{MWage}$ and $T_{FirmMsg}^{LWage}$, we allocate only around 95% and 71% of the total budget to wage, and use the remaining of it to incentivize workers through CSR.

We start our analysis by comparing the rate of accepting the contract across these treatments. As illustrated in Table 7, while the propensity to accept the contract slightly decreases as the wage offer goes down (from 94.92% in $T_{baseline}^{HWage}$ to 94.59% in $T_{FirmMsg}^{MWage}$ and further down to 93.33% in $T_{FirmMsg}^{LWage}$), the declines are not significant (p-values from two-sample test of proportions between pairs of treatments: $P_{2-3} = 0.81$; $P_{2-4} = 0.27$; $P_{3-4} = 0.39$).

Table 7: Acceptance Rates in Treatments 2, 3 and 4

	$T_{Baseline}^{HWage}$ (2)	$T_{FirmMsg}^{MWage}$ (3)	$T_{FirmMsg}^{LWage}$ (4)
Landed on our website and were randomized into a treatment	531	536	540
Accepted the contract	504	507	504
Rate of Acceptance	94.92%	94.59%	93.33%

4.2.1 Share of Cheaters

As wage decreases from \$1.26 to \$1.20 and \$0.90, and is substituted with CSR, more workers cheat. The share of cheaters increases by about 25%, from 0.198 (std. err. = 0.018) in $T_{baseline}^{HWage}$ to 0.247 (std. err. = 0.019) in $T_{FirmMsg}^{MWage}$ (p-value=0.066), and by over 50% from $T_{baseline}^{HWage}$ to 0.300 (std. err. = 0.020) in $T_{FirmMsg}^{LWage}$ (p-value=0.000). The increase in the share of cheaters from $T_{FirmMsg}^{MWage}$ to $T_{FirmMsg}^{LWage}$ is also significant (p-value=0.058).

Table 8: Share of cheaters across treatments 2, 3, and 4

	$T_{Baseline}^{HWage}$ (2)	$T_{FirmMsg}^{MWage}$ (3)	$T_{FirmMsg}^{LWage}$ (4)
Share of Cheaters	0.198 (0.018)	0.247 (0.019)	0.300 (0.020)
Share of Type-1 Cheaters	0.062 (0.011)	0.073 (0.012)	0.097 (0.013)
Share of Type-2 Cheaters	0.147 (0.016)	0.181 (0.017)	0.220 (0.018)

A similar pattern emerges when we look at the number of type-1 and type-2 cheaters, separately. The share of type-1 cheaters increases from 0.062 (std. err. = 0.011) in the $T_{baseline}^{HWage}$ to 0.073 (std. err. = 0.012) in $T_{FirmMsg}^{MWage}$ and to 0.097 in the $T_{FirmMsg}^{LWage}$ (std. err. = 0.013). While the increase in the share of type-1 cheaters from $T_{baseline}^{HWage}$ to $T_{FirmMsg}^{LWage}$ is significant (p-value=0.036), the jumps between $T_{baseline}^{HWage}$ and $T_{FirmMsg}^{MWage}$ (p-value=0.47), and between $T_{FirmMsg}^{MWage}$ and $T_{FirmMsg}^{LWage}$ (p-value=0.167) are not large enough to be significant, given our sample size. The number of type-2 cheaters also increase as we substitute wage for CSR. The share of type-2 cheaters increases from 0.147 (std. err. = 0.016) in the baseline to 0.181 (std. err. = 0.017) and 0.220 (std. err. = 0.018) in $T_{FirmMsg}^{MWage}$ and $T_{FirmMsg}^{LWage}$.²⁴ Figure 9 summarizes these findings.

The conclusions are robust to restricting our sample to workers who completed the task and received full payment (Figure 10). While the propensity to finish the task, conditional on accepting the

²⁴ P-value from $T_{baseline}^{HWage}$ and $T_{FirmMsg}^{MWage}$ =0.137; P-value from $T_{FirmMsg}^{MWage}$ and $T_{FirmMsg}^{LWage}$ =0.124; P-value $T_{baseline}^{HWage}$ to $T_{FirmMsg}^{LWage}$ =0.003.

contract decreases as we substitute wage for CSR from 85.32% in $T_{baseline}^{HWage}$ to 81.85% and 77.98% in $T_{FirmMsg}^{MWage}$ and $T_{FirmMsg}^{LWage}$, the share of cheaters (among those who complete the task) increases from 0.060 (std. err. = 0.012) in $T_{baseline}^{HWage}$, to 0.080 (std. err. = 0.013) and 0.102 (std. err. = 0.015) in $T_{FirmMsg}^{Wage}$ and $T_{FirmMsg}^{LWage}$.²⁵ We summarize these findings in Result-5.

Result-5: Substituting wage with CSR increases cheating on the extensive margin. Reducing workers' wage by 5% and substituting it to CSR increases the share of cheaters by 25%. A further 25% decrease in wage and substitution to CSR, further increases cheating by another 21%.

4.2.2 Intensity of Cheating

Our findings on the extensive margin, are echoed on the intensity of cheating. The intensity of cheating per worker increases by more than 25% when we substitute 5% of the wage with CSR, and it increases by over 53% when we substitute around 29% of the wage with CSR. As illustrated in Figure 11 and Table 9, the average number of images that a worker cheats on, increases from 1.383 (std. err. = 0.140) in $T_{baseline}^{HWage}$, to 1.74 (std. err. = 0.156) and 2.121 (std. err. = 0.168) in $T_{FirmMsg}^{MWage}$ and $T_{FirmMsg}^{LWage}$. The increases in the intensity of cheating from $T_{baseline}^{HWage}$ to $T_{FirmMsg}^{MWage}$ (p-value=0.054), from $T_{FirmMsg}^{MWage}$ to $T_{FirmMsg}^{LWage}$ (p-value=0.066), and from $T_{baseline}^{HWage}$ to $T_{FirmMsg}^{LWage}$ (p-value=0.000) are all significant.

Table 9: Intensity of cheating across treatments 2, 3 and 4

	$T_{Baseline}^{HWage}$ (2)	$T_{FirmMsg}^{MWage}$ (3)	$T_{FirmMsg}^{LWage}$ (4)
Total cheating: average number of images a worker cheated on	1.383 (0.140)	1.740 (0.156)	2.121 (0.168)
Type-1 cheating: average number of readable images a worker misreported as unreadable	0.183 (0.044)	0.162 (0.034)	0.181 (0.033)
Type-2 cheating: average number of images NOT submitted by each worker	1.200 (0.135)	1.578 (0.154)	1.940 (0.168)

The effects on the intensity of cheating, are mainly driven by increases in type-2 cheatings across treatments. While the level of type-1 cheating by all workers does not vary as we substitute wage with CSR across treatments, the average intensity of type-2 cheating increases from 1.200 (std. err= 0.135) images per worker in $T_{baseline}^{HWage}$ to 1.578 (std. err= 0.154) in $T_{FirmMsg}^{MWage}$ and further up to 1.940 (std. err= 0.168) in $T_{FirmMsg}^{LWage}$ (p-values: from $T_{baseline}^{HWage}$ to $T_{FirmMsg}^{MWage}$ =0.101; from $T_{FirmMsg}^{MWage}$ to $T_{FirmMsg}^{LWage}$ =0.129; from $T_{baseline}^{HWage}$ to $T_{FirmMsg}^{LWage}$ =0.002) . Result-6 summarizes our findings on the intensity of misbehavior.

Result-6: Substituting wage with CSR increases the intensity of cheating. Reducing the wage by 5%

²⁵ Given our sample size, only the increase from $T_{baseline}^{HWage}$ to $T_{FirmMsg}^{LWage}$ is significant (p-value=0.029).

and substituting it by CSR increases the average intensity of cheating by over 25%. A further 25% decrease in wage (and substitution with CSR) further increases the intensity of cheating by another 22%.

4.2.3 Quality

Does the quality of work respond to the substitution of wage with CSR? To answer this question, we compare the average Levenstein distance across $T_{baseline}^{HWage}$, $T_{FirmMsg}^{MWage}$ and $T_{FirmMsg}^{LWage}$. The average Levenstein distance between workers' transcription and the actual texts slightly decreases from 23.621 (std. err.= 1.101) in $T_{baseline}^{HWage}$ to 22.795 (std. err.= 0.984) and 22.390 (std. err.= 1.060) in $T_{FirmMsg}^{MWage}$ and $T_{FirmMsg}^{LWage}$. These decreases are all insignificant. The quality of work thus, does not seem to change as a response to the substitution of wage with CSR.

Result-7: Substituting wage with CSR does not significantly affect the quality of transcription by workers.

In summary, given our finding that worker misbehavior increases with CSR, and in line with the findings in previous studies that misbehavior decreases with wage, we find cheating to increase as we substitute wage with CSR. This finding suggests that, given the moral-licensing effect that CSR can generate, an employer who faces a tradeoff between allocating financial resources to motivate workers through private incentive or CSR, and seeks to minimize dishonest behavior on the job would be better off by spending the entire budget on wage.

5 Conclusions

We conducted a natural field experiment with over 3000 workers to study the relationship between CSR and employee misbehavior on the job. We recruited workers through Amazon MTurk and randomized them into treatments where we varied the CSR incentive, CSR framing, and wage. Workers who accepted our work contract were required to complete a short transcription task. The incentive structure was designed in a way that provided opportunities for cheating on the job.

Our empirical analyses suggest that CSR increases worker misbehavior. Workers who received our CSR message were more likely to become cheaters and cheated more often than those who were not incentivized with CSR. In-line with the moral-licensing effect of CSR, (marginally insignificantly) more workers cheated when CSR was framed as a benevolent act on behalf of the worker as opposed to the firm. We also find that substituting wage for CSR significantly increases misbehavior. We find no effects on the quality of work.

An alternative explanation for CSR increases misbehavior may lie in its signaling value. If workers

believe that a firm who engages in CSR is more lenient or less likely to monitor workers' performances, then CSR may increase misbehavior by reducing its cost. While we cannot completely rule-out this potential channel, our finding that (marginally insignificantly) more workers cheat when CSR is framed as a benevolent act on behalf of the worker as opposed to on behalf of the firm provides a suggestive evidence for moral-licensing to be the channel through which CSR increases cheating.

Our findings have important implications for the use of CSR as a tool for motivating workers. Previous studies that have investigated the supply-side effect of CSR have focused on the effect on productivity. Our results suggest that, in an environment where worker misbehavior is of a great concern to the employer, CSR may have counterproductive effects by increasing misbehavior through moral-licensing. Of course, the degree to which CSR can generate moral-licensing depends to a great degree on how it is communicated to the workers. Since moral-licensing works through exploiting image concerns, we believe that framing CSR in a way that triggers worker's self-image may increase the level of misbehavior on the job.

We believe that our findings are likely to be generalizable to the dimensions of a worker's counterproductive behavior that damage their self-image. For CSR to increase misbehavior through moral-licensing, engaging in misbehavior needs to induce some form of guilt feeling in the worker. In other words, an individual does not need a moral-licensing tool (such as CSR) to commit a misbehavior, if that misbehavior does not provoke a feeling of guilt, and does not damage his/her self-image. Therefore, we expect those counterproductive behavior in the workplace that are perceived to be more unethical (such as stealing or lying) to be more likely to increase with CSR.

We recognize a number of limitations on the extent to which our findings can be applied to other settings that we believe future research should seek to address. First, while our design attempted to provide insights on the role of self-selection in the way in which CSR affects misbehavior, due to the very high rate of accepting the offer, we were not able to document any sorting effects. It thus remains an open question whether in environments where the selection effect is more pronounced, firms who engage in CSR attract more honest and ethical workers, or they attract less honest workers who take advantage of the moral-licensing effect of CSR and engage in more misbehavior. Second, our experiment was conducted as a one-shot interaction with contingent workers. While contingent workers shape an increasing portion of the workforce,²⁶ their relationship with the employer, and thus their responses to various policies can be quite different from the ones of an organization's permanent employees. For instance, an employer's relationship with contingent employees is by definition short-lived, and thus reputation concerns may play a much less of an important role, in the choices of contingent workers. The temporary nature of the relationship

²⁶ A report by the U.S. Government Accountability Office estimated that around 40.4% of employed labor force in 2010 was in contingent arrangements a 10 percentage points increase from 2005 (30.6%) (<http://www.gao.gov/assets/670/669899.pdf>).

of employers with contingent employees also affects the extent to which workers identify with the organization that they work for, and consequently with the employer's CSR identity. Thus, the extent to which our findings will generalize in longer-term employment relationships remains to be explored in future research.

6 References

- Association of Certified Fraud Examiners. (1996). *Report to the Nation: Occupational Fraud and Abuse*. Association of Certified Fraud Examiners.
- Backhaus, K. B., Stone, B. A., & Heiner, K. (2002). Exploring the Relationship Between Corporate Social Performance and Employer Attractiveness. *Business & Society*, 41(3), 292-318.
- Bénabou, R., & Tirole, J. (2006). Incentives and Prosocial Behavior (cont.). *American Economic Review*, 96(5), 1652-1678.
- Bénabou, R., & Tirole, J. (2010). Individual and corporate social responsibility. *Economica*, 77(305), 1-19.
- Bénabou, R., & Tirole, J. (2011). Identity, morals, and taboos: Beliefs as assets. *Quarterly Journal of Economics*, 126(2), 805-855.
- 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).
- Blanken, I., van de Ven, N., & Zeelenberg, M. (2015). A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin*, 41(4), 540-558.
- Bradler, C., Dur, R., Neckermann, S., & Non, A. (2016). Employee recognition and performance: A field experiment. *Management Science*, 62(11), 3085-3099.
- Brekke, K. A., & Nyborg, K. (2008). Attracting responsible employees: Green production as labor market screening. *Resource and Energy Economics*, 30(January 2007), 509-526.
- Burbano, V. C. (2015). Getting virtual workers to do more by doing good: Field experimental evidence. Technical report, *Working paper*.
- Burbano, V. C. (2016). Social responsibility messages and worker wage requirements: Field experimental evidence from online labor marketplaces. *Organization Science*, 27(4), 1010-1028.

- Cascio, J., & Plant, E. A. (2015). Prospective moral licensing: Does anticipating doing good later allow you to be bad now?. *Journal of Experimental Social Psychology*, *56*, 110-116.
- Chiou, W. B., Wan, C. S., Wu, W. H., & Lee, K. T. (2011). A randomized experiment to examine unintended consequences of dietary supplement use among daily smokers: taking supplements reduces selfregulation of smoking. *Addiction*, *106*(12), 2221-2228.
- Clot, S., Grolleau, G., & Ibanez, L. (2014). Smug alert! Exploring self-licensing behavior in a cheating game. *Economics Letters*, *123*(2), 191-194.
- Clot, S., Grolleau, G., & Ibanez, L. (2017). Moral self-licencing and social dilemmas: an experimental analysis from a taking game in Madagascar. *Applied Economics*, 1-12.
- Conway, P., & Peetz, J. (2012). When does feeling moral actually make you a better person? Conceptual abstraction moderates whether past moral deeds motivate consistency or compensatory behavior. *Personality and Social Psychology Bulletin*, *38*(7), 907-919.
- Du, S., Bhattacharya, C. B., & Sen, S. (2011). Corporate social responsibility and competitive advantage: Overcoming the trust barrier. *Management Science* (Vol. 57).
- Effron, D. A., Miller, D. T., & Monin, B. (2012). Inventing racist roads not taken: The licensing effect of immoral counterfactual behaviors. *Journal of personality and social psychology*, *103*(6), 916.
- Effron, D. A., Monin, B., & Miller, D. T. (2013). The unhealthy road not taken: Licensing indulgence by exaggerating counterfactual sins. *Journal of Experimental Social Psychology*, *49*(3), 573-578.
- Elfenbein, D. W., Fisman, R., & Mcmanus, B. (2012). Charity as a substitute for reputation: Evidence from an online marketplace. *Review of Economic Studies*, *79*(4), 1441-1468.
- Flory, J. A., Leibbrandt, A., & List, J. A. (2016). The Effects of Wage Contracts on Workplace Misbehaviors: Evidence from a Call Center Natural Field Experiment (No. w22342). *National Bureau of Economic Research*.
- Greening, D. W., & Turban, D. B. (2000). Corporate Social Performance As a Competitive Advantage in Attracting a Quality Workforce. *Business & Society*, *39*(3), 254-280.
- Harrison, G.W. and J.A. List, (2004) Field Experiments, *Journal of Economic Literature*, *42*(4), 1009-1055.

- Hedblom, D., Hickman, B. R., & List, J. A. (2016). Toward an Understanding of Corporate Social Responsibility: Theory and Field Experimental Evidence.
- Jordan, J., Mullen, E., & Murnighan, J. K. (2011). Striving for the moral self: The effects of recalling past moral actions on future moral behavior. *Personality and Social Psychology Bulletin*, *37*(5), 701-713.
- Kouchaki, M., & Jami, A. (2016). Everything We Do, You Do: The Licensing Effect of Prosocial Marketing Messages on Consumer Behavior. *Management Science*.
- Kosfeld, M., & Neckermann, S. (2011). Getting more work for nothing? Symbolic awards and worker performance. *American Economic Journal: Microeconomics*, *3*(3), 86-99.
- KPMG, (2015). KPMG International survey of corporate responsibility reporting 2015. Amsterdam, The Netherlands: KPMG.
- 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.
- Lii, Y. S., & Lee, M. (2012). Doing right leads to doing well: When the type of CSR and reputation interact to affect consumer evaluations of the firm. *Journal of business ethics*, *105*(1), 69-81.
- 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.
- Mazar, N., & Zhong, C. B. (2010). Do green products make us better people?. *Psychological science*, *21*(4), 494-498.
- Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral Self-Licensing: When Being Good Frees Us to Be Bad Empirical Demonstrations of Moral Self-Licensing. *Social and Personality Psychology Compass*, *4*(5), 344-357.
- NRF, (2016). National Retail Security Survey 2016. Report, Department of Sociology and the Center for Studies in Criminology and Law, University of Florida.
- Ockenfels, A., Sliwka, D., & Werner, P. (2015). Timing of kindness - Evidence from a field experiment. *Journal of Economic Behavior and Organization*, *111*, 79-87.
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*(5).

- 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.
- Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological science*, *20*(4), 523-528.
- Sen, S., Bhattacharya, C. B., & Korschun, D. (2006). The role of corporate social responsibility in strengthening multiple stakeholder relationships: A field experiment. *Journal of the Academy of Marketing Science*, *34*(2), 158-166.
- Servaes, H., & Tamayo, A. (2013). The impact of corporate social responsibility on firm value: The role of customer awareness. *Management Science*, *59*(5), 1045-1061.
- Tiefenbeck, V., Staake, T., Roth, K., & Sachs, O. (2013). For better or for worse? Empirical evidence of moral licensing in a behavioral energy conservation campaign. *Energy Policy*, *57*, 160-171.
- Tonin, M., & Vlassopoulos, M. (2015). Corporate Philanthropy and Productivity: Evidence from an Online Real Effort Experiment. *Management Science*, *61*(8), 1795-1811.
- Turban, D. B., & Greening, D. W. (1997). Corporate Social Performance and Organizational Attractiveness To Prospective Employees. *Academy of Management Journal*, *40*(3), 658-672.
- Urbszat, D., Herman, C. P., & Polivy, J. (2002). Eat, drink, and be merry, for tomorrow we diet: Effects of anticipated deprivation on food intake in restrained and unrestrained eaters. *Journal of abnormal psychology*, *111*(2), 396.

7 Figures

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.
 - o 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

Figure 1: A snapshot of the Human Intelligence Task (HIT) on Amazon MTurk

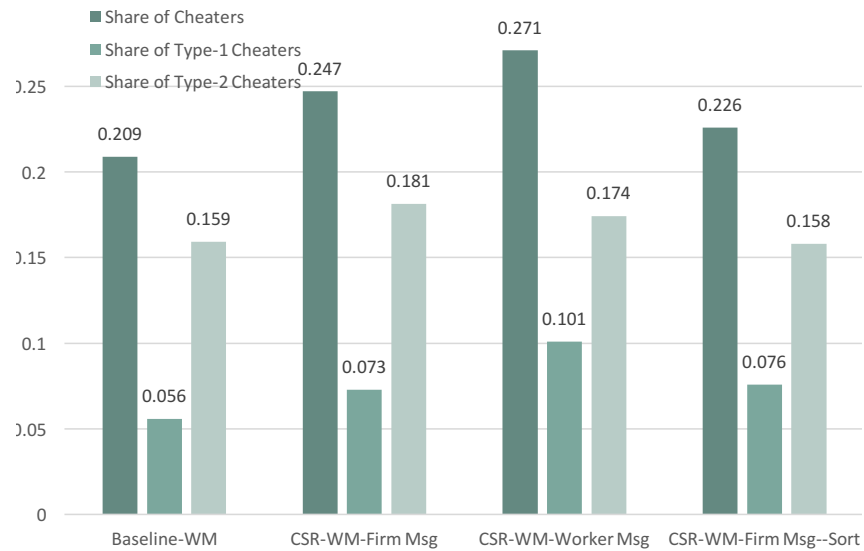


Figure 2: Share of cheaters in treatments 1, 3, 5 and 6.

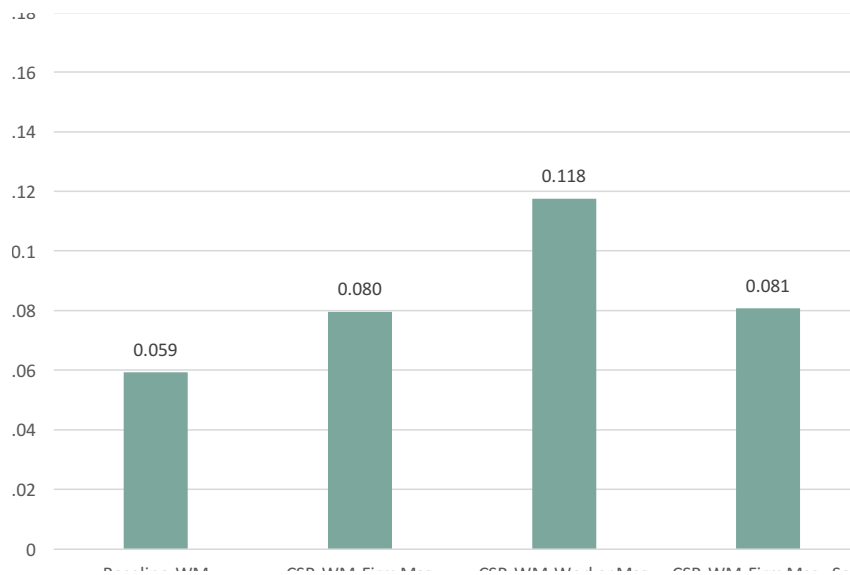


Figure 3: Share of cheaters among workers who completed the task treatments 1, 3, 5 and 6.

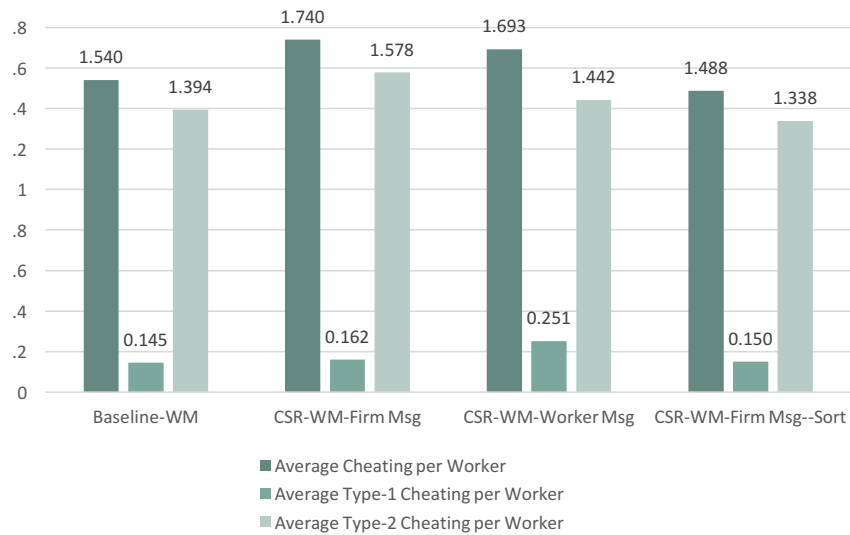


Figure 4: Average number of images a worker cheated on in treatments 1, 3, 5 and 6.

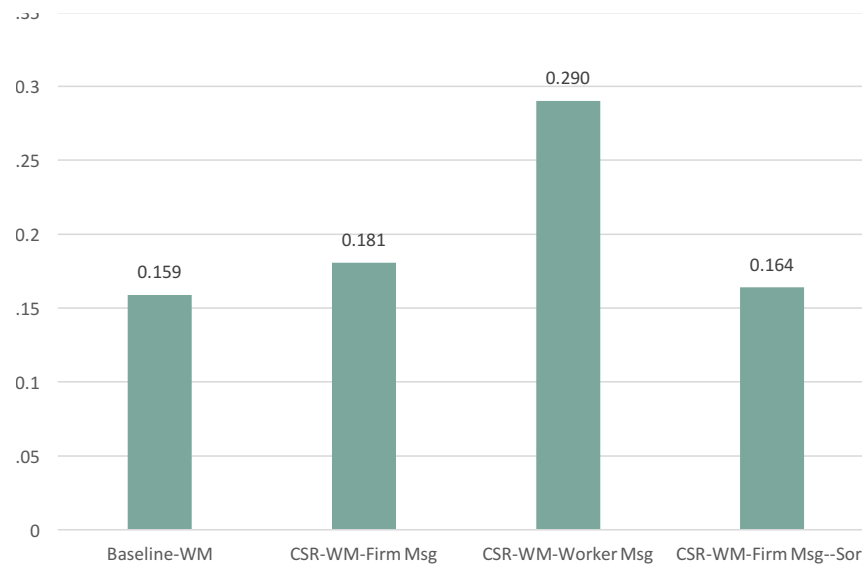


Figure 5: Average number of images a worker cheated on among those who completed the task in treatments 1, 3, 5 and 6.

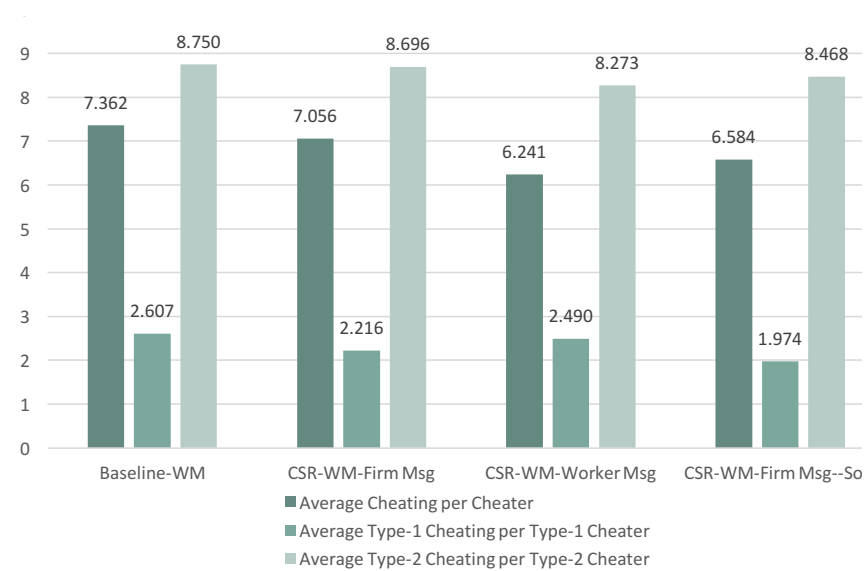


Figure 6: Average number of images a cheater cheated on, in treatments 1, 3, 5 and 6.



Figure 7: Fraction of cheaters who cheat at different levels, across baseline and CSR treatments with no sorting.

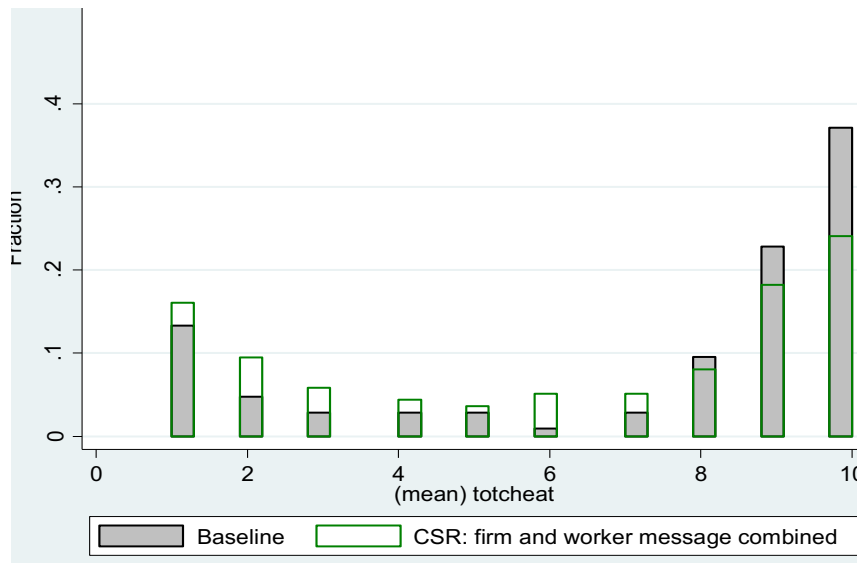


Figure 8: Fraction of cheaters who cheat at different levels, across treatments 1 and 5.

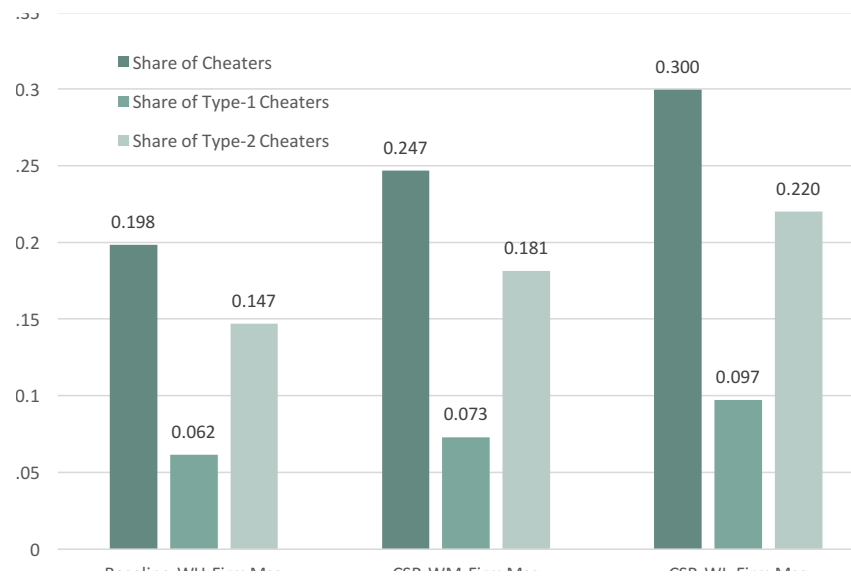


Figure 9: Share of cheaters in treatments 2, 3 and 4.

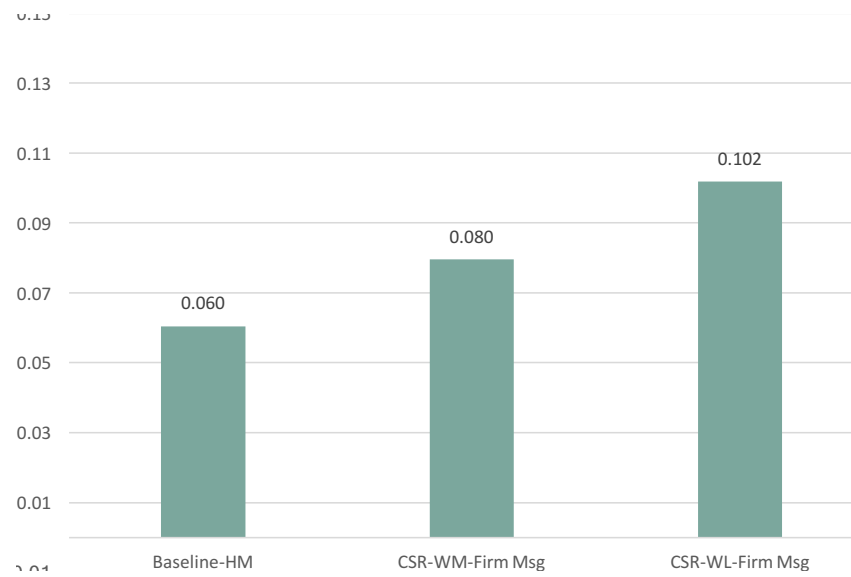


Figure 10: Share of cheaters among workers who completed the task treatments 2, 3 and 4.

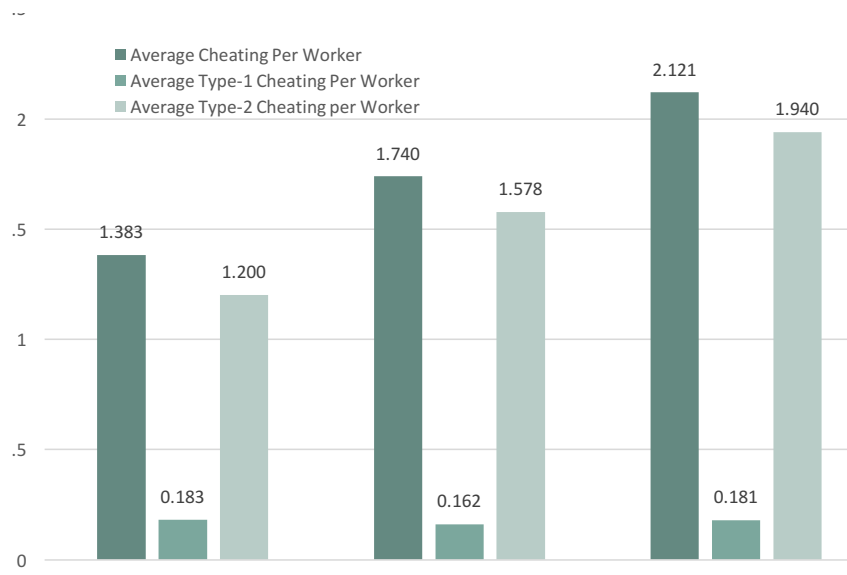


Figure 11: Average number of images a worker cheated on in treatments 2, 3 and 4.