I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12905

When Too Good Is Too Much: Social Incentives and Job Selection

Tommaso Reggiani
Rainer Michael Rilke

JANUARY 2020

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12905

When Too Good Is Too Much: Social Incentives and Job Selection

Tommaso Reggiani
Cardiff University, Masaryk University and IZA

Rainer Michael Rilke
WHU-Otto Beisheim School of Management

JANUARY 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

When Too Good Is Too Much: Social Incentives and Job Selection*

We analyze the effects of substitutability of social incentives on the labor supply of gigworkers (N=944) in a natural field experiment. In our treatments, we vary the proportion of the worker's wage that is donated to a social cause. Our experimental design allows us to observe the decision to accept a job (extensive margin) and different dimensions of productivity (intensive margin). The results show that when the worker has to donate small or moderate parts to a prosocial organization, labor supply on the extensive margin remains unaffected, but productivity on the intensive margin increases; when workers have to give larger portions of their wages, labor supply and productivity decrease. When workers have to donate parts of their wages to an antisocial cause, labor supply on the extensive and intensive margin is negatively affected. We discuss the implications of these results for the understanding of social incentives and corporate social responsibility on labor supply.

JEL Classification: C93, D23, M52

Keywords: social incentives, labor supply, CSR, field experiment

Corresponding author:
Tommaso Reggiani
Masaryk University
Faculty of Economics and Administration
Department of Public Economics
Lipová 507/41a - 602 00 Brno
Czech Republic
E-mail: tommaso.reggiani@econ.muni.cz

* We thank Milos Fisar, Caterina Gianetti, Bernd Irlenbusch, Ondrej Krcal, Felix Kölle, Jiri Spalek, Rostislav Stanek and participants of the Workshop for Experimental Economics at the Max-Planck Institute (EEG) Bonn 2019 and participants of the Forschungskolloquium at the WHU for helpful comments and suggestions; Martin Both, Petr Hanzlik, and Alessia Sconti for excellent research assistance. Financial support from the Czech Science Foundation (GACR grant n. GA20-06785S) is gratefully acknowledged.

1 Introduction

Misaligned incentive systems of financial institutions were one reason that contributed to the financial turmoils of 2008. In the aftermath, banks were forced to rethink their remuneration systems and reconcile incentives with societal needs (Dewatripont et al., 2015). One example of how incentives could be aligned with societal needs is the bank Credit Suisse. In 2011, the bank substituted part of its employees' wages with donations to charitable causes. In particular, its high-level employees in the US were supposed to donate 2.5 percent of their yearly bonus to a charitable organization (Financial Times, 2011). The bank estimated that a total of roughly $30 million would be raised for prosocial activities.[1]

With this study, we aim at contributing to the pressing research agenda of how corporate social responsibility influences labor supply. We are focusing on an essential element of corporate social responsibility: social incentives, i.e., incentives defined as donations to causes not linked to the organization's core objectives (like in the case of Credit Suisse) and investigate its role on labor supply and effort provision of workers.

Consider a person looking for a job. One firm offers her a contract with a specific wage, while an alternative firm offers her a contract with a lower wage, but additionally states that a part of the profits generated through her efforts is donated to a social cause.

In this paper, we ask: Can an employer lower the wage and substitute it with a social incentive to make her accept the contract? Given that she agrees to a lower personal wage, will this influence her effort provision on the job? Moreover, does it depend on the prosociality of the supported cause? The above example illustrates that contracts involving social incentives can be designed differently with respect to the relative size and perceived prosociality of the social incentive.[2]

Donating profits to organizations not linked to the company's main objectives has been subject to criticism since Milton Friedman's famous critique of firms' corporate social responsibility activities and philanthropic activities of managers (Baron, 2007; Friedman, 1970). Despite questions about the moral duty of managers to maximize profits (Gond and Nyberg, 2017), the effect of the company's prosocial activities on employees' motivation has attracted only limited attention by OS scholars and managerial economists (Cassar and Meier, 2018).

Corporate donations might generate differential motivational effects on employees. Onthe one hand, the money spent on donations is not given to the employees and, in turn, can lead to lower performance. On the other hand, if the supported cause aligns with the employees' values, it might enhance their motivation and performance.

We approach our research questions by running a natural field experiment in an online labor market. We design a field experiment to observe when workers accept to work (labor supply) for a contract that specifies social incentives, and how much effort they are willing to spend on the

[1] Not only banks and corporations spend money on charitable causes and make it public (Bagnoli and Watts, 2017; Calveras and Ganuza, 2016); recent numbers by the Charities Aid Foundation (2018) report that FTSE 100 companies donated £1.9b in 2016, which accounts for about 2.5 percent of their pre-tax profits.

[2] A survey by America's Charities (2017) among American workers yielded that workers want to work for companies that reinforce their values. More than 70 percent of respondents said it was imperative or essential to work for an employer where mission and personal values align. No trust in the charitable cause was the main reason for employees to be critical about the company's prosocial activities.

job in terms of quantity (measured as the number of output they produce), speed (measured as time they need to fulfill the task), and work quality (measured as how accurately they perform the task).

First, we hire workers to answer a lengthy survey for a fixed payment. Second, when finished with this assignment, workers are asked to perform an additional real-effort task for a bonus. This second activity constitutes our core experimental task.

In our control treatment (*0%*), the entire bonus is given to the worker, and nothing is donated. In our treatment variations, we increase the amount that is donated and deducted from the worker's wage. In treatments *10%*, *50%*, *90%*, and *100%*, ten, fifty, ninety, and a hundred percent of the bonus is donated, respectively.

To understand how workers respond to the kind of charitable cause, we vary the beneficiary of the donation. In one set of experimental treatments, the donation is given to a prosocial cause (United Nations Children's Emergency Fund, UNICEF), whereas in another set of experimental treatments it is given to an antisocial cause (National Rifle Association, NRA).[3]

Surprisingly, reducing the wage a little bit (in treatments *10%* and *50%*) and donating this reduction to a prosocial cause does not lead to a lower labor supply compared to our control treatment where workers receive the complete bonus. Once on the job, these workers are significantly more productive compared to workers in the control treatment. Once the prosocial donation becomes too large (in treatments *90%* and *100%*), labor supply drops, and productivity is negatively affected when the entire bonus is donated (treatment *100%*). For antisocial incentives, even small donations negatively affect labor supply. Under antisocial incentives, workers tend to work more slowly and less accurately.

Our study contributes to a stream of literature that focuses on the supply side of corporate social responsibility and prosocial incentives. This research, most notably initiated by Ariely et al. (2009), studies the circumstances under which participants of laboratory experiments are willing to exert effort when their wages are donated.[4] Prior laboratory experiments focused their attention on the effect of prosocial incentives on the level of effort (intensive margin) exerted by the workers. Selection effects, i.e., how workers accept contracts with or without social incentives, are particularly hard to study in a lab setting, where they — in principle — have to leave the laboratory without payment when they would reject the offer. In the laboratory, monitoring of the experiment, peer-effect, and possible demand effects can can influence subjects' decision to leave the lab and might therefore deliver results with limited external validity. A newer stream of papers focuses on settings with presumably higher external validity due to lower costs of rejecting offered contracts: online gig labor markets.

A series of studies documents a positive effect of informing gig workers about the prosocial activities of the employer (Burbano, 2016, 2019), and explaining why efforts create a positive externality for a charitable organization (Cassar and Meier, 2017) on effort provision.[5]

[3]These organizations have been used in prior experiments to operationalize prosocial and antisocial incentives (Ariely et al., 2009; Kajackaite, 2015).

[4]More examples of these experiments can be found in Charness et al. (2016); Conrads et al. (2016); Imas (2014); Kajackaite and Sliwka (2017); Tonin and Vlassopoulos (2015), generally documenting a positive effect of prosocial incentives.

[5]Although these studies do not explicitly study selection effects, an indication for their existence can be found in Cassar and Meier (2017). Here it is reported that when a prosocial incentive is offered, almost 10.8 percent

Most closely related to the study at hand are the field experiments of Hedblom et al. (2019) and Schwartz et al. (2019, Study 2). Both study the influence of CSR on job selection in field experiments in two distinct ways. While the former manipulate the information job a candidate receives about the CSR activities of the employer, the latter contrast treatments in which job candidates are either paid individually or not paid. However, their entire wage is given to a charitable organization (resembling our *100%* treatment). Hedblom et al. (2019) show that significantly more workers choose to work when the company stresses their prosocial activities, whereas Schwartz et al. (2019) observe that the number of workers is lower when their wage is donated.

One reason for these seemingly contradictory results lies in the difference in the compensation of workers. While While the workers in Schwartz et al. (2019) have to bear the entire costs of prosociality by giving up their wages, which leads to a reduction in labor supply, in Hedblom et al. (2019) prosociality comes as an additional benefit to the workers' wages, which leads to an increase in labor supply. In both cases, workers who accept the job exert more effort, which can be explained by a selection of some motivated workers.

While it is true that, in firms, workers never give up their entire wage for a prosocial cause, corporate social responsibility is also not costless and comes at the expense of someone in the organization.

In contrast to prior field experiments, in our study, we vary how much of the cost of corporate social responsibility workers have to bear. We systematically compare different wage/donation combinations and their effect on labor supply and effort provision. In addition to that, we investigate cases where the social cause cannot be seen as clearly prosocial and in line with the worker's values, but is somewhat antisocial and potentially orthogonal to the workers' values.[6] Both questions are are highly relevant when it comes to designers of incentives and remuneration systems planning to introduce social incentives.

2 Experimental design

2.1 General setup and subject pool

In our natural field experiment, we hire workers to take part in a composite survey on the online labor platform Amazon Mechanical Turk (mTurk). mTurk offers the opportunity to restrict the job posting to specific subgroups. For this study, we limit our sample to workers located in the United States with more than 50 approved jobs in the past and an approval rate of more than 75 percent. Thus, workers eligible for the HIT were somewhat experienced with online jobs and with the platform and its procedures.[7] mTurk has been subject to many new studies by

(144/1332) of the workers do not finish the task and drop out. This figure is only 7.9 percent(107/1338) when no prosocial incentive is offered (p=.013, χ^2-test). Selection effects helping incentives are also documented in Ashraf et al. (2018). Related are also Burbano and Chiles (2018) and List and Momeni (2017) who investigate the role of employer CSR on worker misconduct and find mixed evidence.

[6]Somewhat related is the work by Carpenter and Gong (2016). They study how the effect of job missions is moderated by a piece-rate incentive. When workers are paid a fixed wage, they provide more labor then their values are in line with the job's mission. Introducing piece-rate incentives leads workers to exert more effort and makes up a large share of the loss in output due to mismatching.

[7]Jobs posted on mTurk are called HITs (Human Intelligence Tasks). Our HIT was announced as a "Survey on personality". In order not to attract workers specifically interested in the subject of prosocial incentives or

economists. Goodman et al. (2013) shows that mTurk workers do not behave any differently from other other samples and differ little on any other dimension. Gig workers for our research are hired to fill multiple lengthy surveys.

Within the description of the HIT, we announced that subjects would take about 15 minutes to complete the survey. Workers who wanted participate in our HIT were redirected to a homepage. This website informed them about surveys on personality. For completing all surveys, workers receive a fixed payment of \$.50.[8]

In total, we administered a series of four surveys: i) a self-reported altruism assessment scale (Rushton et al., 1981); ii) a domain-specific risk-taking questionnaire (Blais and Weber, 2006); iii) the Machiavellian personality scale (Christie and Geis, 1970); iv) questions on gender and age. Surveys were presented in randomized order to prevent potential spillover-effects. An overview of the sequence of all events is displayed in Table 1.

Self-reported altruism is elicited via the scale proposed by Rushton et al. (1981). It comprises a battery of 20 items in which respondents have to rate the frequency with which they have engaged in altruistic behavior on a 5-point Likert scale (1=never; 5=very often). The degree of altruism is aggregated into an individual score ranging from 20 (not altruistic) to 100 (extremely altruistic).

Finally, we gathered information about the individual degree of Machiavellianism . This is a personality trait that captures how much a person is prone to harm third parties to achieve a private gain or goal. It is elicited with twenty items describing ways of thinking and opinions about people on a 6-point Likert scale (1=strongly disagree; 6=strongly agree) (Christie and Geis, 1970).

Individual gradients of risk sensitivity, both in financial and ethical domains, are measured with a 5-item battery mutated by Blais and Weber (2006). These two context-specific risk scores are obtained, averaging multiple self-reported ratings on a 7-point Likert scale (1=very risk-averse; 7=very risk-loving).

We planned to stop the HIT once we had collected in expectation 100 observations per treatment; when we reached 900 workers, we closed the HIT, but still had gig workers doing the survey; thus, we ended up with 944 workers. In total, we needed one day to collect the respective number of observations. Sixty percent of workers in our experiment were female, with an average age of 38.1. On average, they needed 14 minutes to complete the study (survey + experiment).

2.2 Treatments and experimental task

Once the workers had finished the survey, we informed them that there was an additional job to do. For this additional job, they would receive a bonus B that would be added to the fixed payment of \$.50. In the instructions for the additional job, we stressed that it was entirely voluntary to do the additional job and that they could freely choose to skip the job and end the study without negative consequences for the approval of their payments. When workers

something similar, we choose this rather neutral title.

[8]This corresponds to an hourly wage of \$4, which is above the median hourly wage on the platform (Hara et al., 2018).

Table 1: Sequence of events in the experiment

Step	Content	Measures
1	Survey measurement	Altriusm, Machiavellianism, Risk, Demographics
↻	Treatment randomization (see Table 2)	
2	Decision about additional job	Extensive margin (Accept or reject job offer)
3	Real effort task (Figure A.1)	Intensive margins (Quantity, quality, completion time)
4	Payments realized	-

rejected the offer, however, no bonus was paid, and no donation made. In addition to this, the instructions explained the real-effort task. We asked the workers to look at ten pictures and to find keywords that could be used to describe the picture.[9] Further, we explained that they should find as many keywords as they could, so that we would be able to interpret a higher quantity as a more favorable reaction to the principal's request.

After these instructions – on the same screen – the bonus was explained. Workers were

Table 2: Treatments, incentives, and observations.

Treatment	Total bonus B	Amount of B given to ... ... worker	... social cause	Obs.	Age	Female
0%	\$ .50	\$ .50	\$ 0	111	37.23	.61
UNICEF						
10%	\$ .50	\$ .45	\$ .05	101	35.60	.50
50%	\$ .50	\$ .25	\$ .25	103	38.00	.71
90%	\$ .50	\$ .05	\$ .45	113	38.58	.55
100%	\$ .50	\$ 0	\$ .50	108	36.74	.68
NRA						
10%	\$ .50	\$ .45	\$ .05	107	37.92	.55
50%	\$ .50	\$ .25	\$ .25	100	38.54	.58
90%	\$ .50	\$ .05	\$ .45	103	41.90	.57
100%	\$ .50	\$ 0	\$ .50	98	38.46	.63

Notes: The table displays the description of nine treatments with the respective worker bonus/donation configuration. Obs., age, and female shows the number of observations, the average worker's age, and the relative frequency of female workers per treatment, respectively.

[9] One example is shown in Figure A.1 in the Appendix. We chose this real-effort task because it a common task in online labor markets (see Kässi and Lehdonvirta, 2018). Results from these image-tagging exercises are used to feed algorithms for image-recognition purposes. A standard laboratory real-effort task (e.g., the slider task/encryption task, Gill and Prowse, 2012; Gall et al., 2019) would have been perceived as a rather artificial artificial HIT hosted by mTurk).

randomly assigned to one out of nine treatments. In our experiment, we varied the proportion of the worker's wage that was donated. In Table 2, we present an overview of all treatments. Our baseline treatment is called *0%*. This number indicates that *0%* of the bonus is donated, and the worker receives $.50 as an additional payment. In the experimental treatments, we introduce a social incentive. In the treatment *10%*, 5 cents are donated, and the worker receives 45 cents as an additional payment. The treatments *50%*, *90%*, and *100%* are designed accordingly.[10]

In addition to the bonus/donation combination, we vary whether the social incentive is prosocial or antisocial, i.e., the beneficiary of the donation. The donation is either made for UNICEF, which can be considered as prosocial or to the NRA, which most subjects are likely to deem as antisocial.[11]

Please note that in the *0%* treatment — where no donation is made — neither UNICEF nor NRA are mentioned. After reading the instructions and being informed about the bonus, workers were asked whether or not they wanted to perform the additional task. When they rejected the offer, they were redirected to the end of the survey. When they accepted the job, the task started.

From a plain economic perspective, assuming myopic, self-regarding, purely money-maximizing preferences we can expect that workers accept the additional job in all treatments and provide no effort.[12]

Outcome-based models of inequity aversion (e.g., Fehr and Schmidt, 1999) might explain why workers reject job offers. Our treatments create inequity in wages between the worker and the social cause. It can be shown that for plausible assumptions on the parameters of the utility function we should observe workers rejecting the job offer when the donation becomes larger than the workers' wage (in treatments *90%* and *100%*).[13]

However, these models are silent about how workers perceive the other entity and compare themselves to the social cause. In fact, according to this reasoning, the beneficiary of the donation should not influence workers' acceptance between prosocial and antisocial incentives on the extensive margin. If instead we consider workers' perceived prosociality of the social cause to trigger inequity aversion, we should observe that more workers give their acceptance for prosocial than for antisocial incentives.

3 Results

We present our results around our main research questions: How far can social incentives be a substitute for monetary incentives? How do they influence labor supply and subsequent effort?

[10] To increase the credibility of this statement, we provided subjects with a link to the personal website of the researchers, where we published the total amount donated one week after the end of the study.

[11] After the experiment we asked all gig workers *How much do you trust UNICEF/NRA?* (1= not at all; 5= very much) and observe that gig workers trusted UNICEF (3.3) significantly more than NRA (2.55) ($p < .0001$, MWU test); which is also in line with the existing literature on social incentives (Ariely et al., 2009; Kajackaite, 2015). Thus, we can safely assume that both organizations are perceived differently in terms of their prosociality. One might also argue that our subjects have different knowledge about the NGOs. When asking *"Did you know UNICEF/NRA prior to the experiment?"* 82 percent (84 percent) of our workers indicated that they knew UNICEF (NRA) ($p= .512$, χ^2 test).

[12] In the treatment *100%*, a selfish, myopic, and purely money-maximizing worker is indifferent between accepting and rejecting the offer. She would surely not exert any effort.

[13] We compute the predictions in Table A.1 of the Appendix.

3.1 Data analysis

We develop our analysis by employing a regression framework that allows us to take into consideration a rich array of idiosyncratic characteristics that could influence such a decision. For this purpose, we complement treatment dummies with control variables elicited in three questionnaires workers answered before performing the core real-effort task.[14]

Our empirical strategy can be summarized in the following model:

$$\begin{aligned} y_i = {} & \beta_0(0\%) \\ & + \beta_1(10\%_{\textbf{UNICEF}}) + \beta_2(50\%_{\textbf{UNICEF}}) + \beta_3(90\%_{\textbf{UNICEF}}) + \beta_4(100\%_{\textbf{UNICEF}}) \\ & + \beta_5(10\%_{\textbf{NRA}}) + \beta_6(50\%_{\textbf{NRA}}) + \beta_7(90\%_{\textbf{NRA}}) + \beta_8(100\%_{\textbf{NRA}}) \\ & + \beta_n[Controls] + \epsilon \end{aligned} \tag{1}$$

We consider the following dependent variables (y_i) for our estimation: In Section 3.2, we start the analysis, addressing the extensive margin investigating the impact of the different intensities of social incentives on *labor supply* (L), measured as the share of the workers accept the job.

In a second step, we address the intensive margin, namely effort provision provided by the gig workers. We analyze the effort provision from a quantitative and and from a qualitative perspective.

Section 3.3 focuses on a simple measure of effort provision: The *quantity* (Q) of keywords provided by each gig worker who decided to accept the job.

Section 3.4 addresses a complementary measure of effort provision that consists of the *completion time* (T), namely the time workers took to process the ten assigned pictures. In this case, the outcome variable is the time (in seconds) workers took to provide keywords for the ten pictures.

Moving from a quantitative to a more qualitative notion of effort provision, Section 3.5 hosts a specific analysis that focuses on the level of *accuracy* (A) of keywords provided by each worker. We implemented the qualitative rating, letting independent raters evaluate the accuracy of all the keywords that had been submitted by the gig workers. For each of the ten pictures, we compiled a list of all the keywords provided. Two research assistants received these lists with the respective picture. After we explained to them the gig workers' task, they were asked to evaluate, on a 5-point Likert scale (0=not accurate, 4=very accurate), how the particular keyword was accurate in describing the picture. The two research assistants (one female and one male) were blind to the treatments and research questions. They rated the answers independently of each other. We asked raters to rate keywords that matched the picture, but contained typos, with the lowest possible rating.[15] The ratings of the keywords were matched back to the gig workers answers;

[14] Since all measures of the questionnaires adopt different metrics, in our regressions we will consider their z-standardization to allow for relative comparisons. Due to the fact that our data-generation process is based on treatment variations that are randomized among a random set of gig workers, and all the treatment outcomes can be contrasted against the actually observed counterfactual control condition, Heckman (1976) latent selection caveats do not directly apply to our case. As conventional for this vein of literature (e.g., Ashraf et al., 2018; Conrads et al., 2016) we rely on parsimonious and easily interpretable OLS and LPM regression models. Qualitatively comparable results, available upon request, are delivered by non-linear estimations.

[15] We observed high inter-rater consistency between the two raters (Cronbach's $\alpha = .89$). Charness and Grieco (2018) implemented a similar rating protocol.

for each worker, we computed the average rating as an indicator for their provided work quality. Our procedure ensured that each keyword received the same rating, independently of the group of keywords they appear in (i.e., a specific keyword received the same rating separate of the treatment). Moreover, ratings are independent of quantity. There is no significant correlation between quantity and quality (Spearman's ρ =-.06, $p = .186$).

Finally, Section 3.6 concludes our analysis. Here we investigate the impact of the different levels of social incentives on a more holistic measure of work efficiency. The idea is to combine both quantitative and qualitative performance (Sauermann, 2016). For this purpose, we construct a multidimensional *work efficiency index*:

$$E = z\left(Quantity \cdot \frac{Accuracy}{Completion\ Time}\right),$$

in which the crude measure of productivity is weighted both by the gradient of accuracy of the delivered output and the time that workers took to process the task. This metric allows us to condense – in one single index – the complete set of multifaceted outcomes we can observe in our experiment. This measure of efficiency is particularly informative, since it is sensitive to the speed-accuracy tradeoff (Förster et al., 2003) that characterize effort provision in multidimensional tasks (e.g., Laske and Schroeder, 2017; Charness et al., 2018; Hedblom et al., 2019).

3.2 Extensive margin: Labor supply

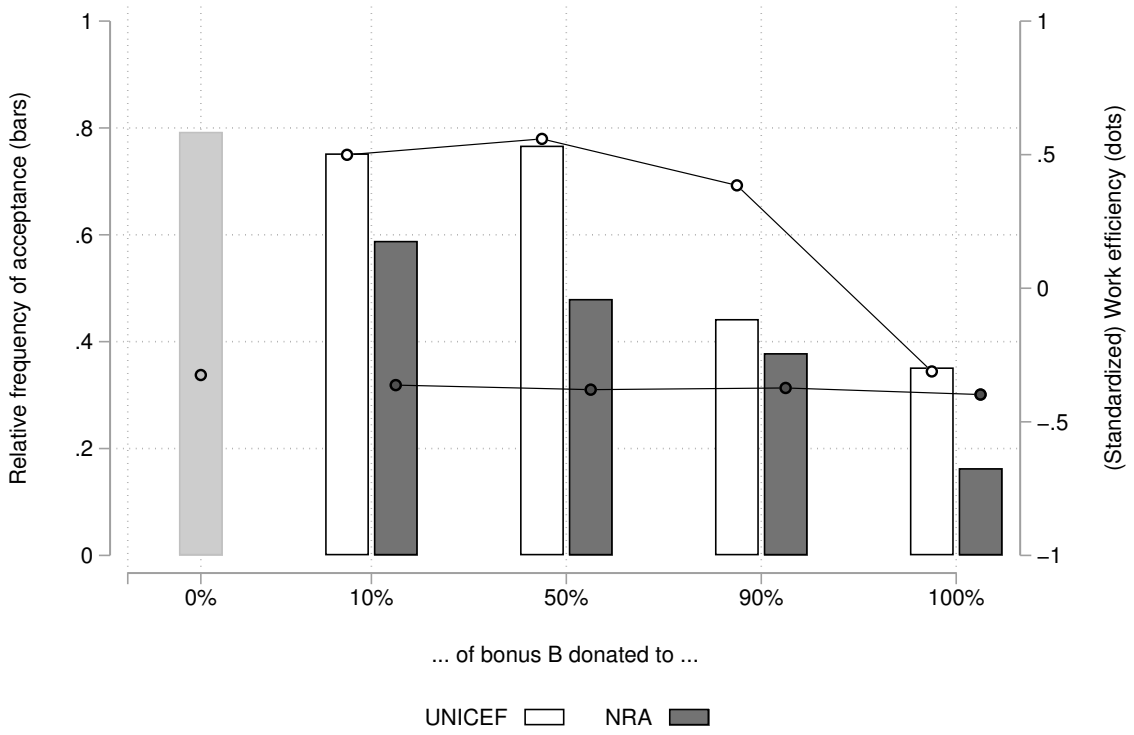
We observe significant differences in labor supply, measured as the number of workers who accepted the additional task between the treatments. The bars in figure 1 display the relative frequency of workers accepting the contract. As visual inspection of the gray bar already reveals that the relative frequency of accepting the job is .79 for the treatment *0%*!, when the bonus is given entirely to the worker.

In treatments *10%* and *50%*, where part of the bonus is given to UNICEF, we observe a relative frequency of .75 and .77, respectively. These differences are not significantly different ($p = .484$, $p = .649$; χ^2-test, respectively). Labor supply drops when the donation to UNICEF is larger than the bonus given to the worker in treatments *90%* and *100%*. A fraction of only .44 and .35 accept the job, respectively. Compared to the *0%* treatment this is a statically significant difference ($p < .01$ and $p < .01$, respectively; χ^2-test).

Acceptance rates are different for the NRA and decline monotonically, compared to the *0%*! treatment. While we observe a fraction of .59 workers accepting the job when *10%* are donated, this figure drops to .16 when the bonus is entirely given to the NRA (treatment *100%*). Comparing the relative frequency of workers accepting the job when parts of the bonus are given to the NRA to the *0%* treatment yields statistically significant differences for all bonus-donation combinations ($p < .01$; χ^2-test).

A linear model (Table 3) predicts the probability to accept the job (labor supply (L)) between our treatments. It allows us to take into account an array of idiosyncratic traits that could influence acceptance. As covariates, we use the measures from three main questionnaires subjects answered in the first step of the experiment. Additionally, we control for gender and age.

Figure 1: Labor supply, effort provision, and social incentives



Notes: The bars show the relative frequency of workers who accepted the additional job (extensive margin) across the different treatments. The dots indicate the average number of words provided by workers who accept the job (intensive margin). White (dark grey) elements indicate the relative frequency for the UNICEF (NRA) treatments. The grey bar and dot show the behavior when no charitable organization is involved.

Table 3: Labor supply (L) and social incentives

	Dependent variable: Acceptance			
Independent variables	(1)	(2)	(3) -males-	(4) -females-
UNICEF				
10%	-0.040	-0.024	-0.056	0.018
	(0.058)	(0.058)	(0.096)	(0.071)
50%	-0.026	-0.041	-0.029	-0.046
	(0.057)	(0.057)	(0.110)	(0.067)
90%	-0.350***	-0.345***	-0.511***	-0.207***
	(0.061)	(0.059)	(0.091)	(0.077)
100%	-0.441***	-0.443***	-0.520***	-0.404***
	(0.060)	(0.059)	(0.099)	(0.073)
NRA				
10%	-0.204***	-0.201***	-0.104	-0.283***
	(0.061)	(0.063)	(0.099)	(0.082)
50%	-0.313***	-0.320***	-0.348***	-0.288***
	(0.063)	(0.062)	(0.102)	(0.079)
90%	-0.414***	-0.413***	-0.411***	-0.408***
	(0.062)	(0.060)	(0.098)	(0.078)
100%	-0.630***	-0.644***	-0.540***	-0.706***
	(0.054)	(0.054)	(0.094)	(0.065)
Control variables				
Female		0.140***		
		(0.032)		
Age		-0.001	0.001	-0.002
		(0.001)	(0.002)	(0.002)
Altruism		0.072***	0.070***	0.081***
		(0.016)	(0.026)	(0.019)
Risk: financial		0.016	0.035	0.004
		(0.019)	(0.031)	(0.025)
Risk: ethical		-0.010	-0.043	0.021
		(0.022)	(0.034)	(0.029)
Machiavellianism		-0.000	0.007	-0.005
		(0.017)	(0.028)	(0.021)
0%(Const.)	0.793***	0.745***	0.706***	0.919***
	(0.039)	(0.064)	(0.099)	(0.077)
N	944	944	379	565
R^2	0.165	0.205	0.204	0.223

Notes: Coefficients (robust standard errors in parentheses) of a linear probability model predicting labor supply. Dependent variables: 1 if worker accepts the additional job, 0 if not. Altruism, Risk: financial, Risk: ethical, and Machiavellianism are z-standardized. Significance indicators: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Model (1) only includes dummy variables for the different treatments (with *0%* being the reference category). Similar to the non-parametric analysis from above, moderate shares of prosocial donations (UNICEF *10%* and *50%*) do not negatively impact the workers' willingness to accept the job. Model (2) refines the analysis controlling both for demographics as well as for the elicited individual traits. Treatment coefficients are stable and treatment coefficients are not affected by the introduction of the controls in the Model (2). Overall, female workers appear to be more willing to accept the additional job (14 percentage points), while age does not play any substantial role.

Workers reporting a higher degree of altruism are systematically more likely to accept the job (a one standard deviation increase in the score positively affects the probability of accepting to perform the task by seven percentage points). Risk sensitivities and Machiavellianism do not turn out to be statistically significant and are negligible in their sizes.

Since the coefficient for female workers is highly statistically significant and economically meaningful in its magnitude, we replicate the same analysis, splitting the sample for males (Model 3) and females (Model 4), respectively. In the *0%* treatment, male workers tend to accept the job less frequently (70 percent) than females (92 percent). This analysis reveals a positive effect of female workers when exposed to high levels of prosocial donations, UNICEF *90%* and *100%*. While for male workers – in both cases – the acceptance rate decreases by approximately -72 percent,[16] the corresponding effects for the female sample are milder and amount to -22 percent and -44 percent, respectively.

Result 1 (Labor supply and prosocial incentives) *Compared to treatment 0%, prosocial incentives attract the same number of workers when 10% or 50% of the bonus is donated. The number of workers decreases when 90% or 100% is donated.*

The negative and significant coefficients for all antisocial bonus-donation combinations highlight the negative impact of antisocial incentives on labor supply (Models 1 and 2). Interestingly, in Model (3), we can observe how a small share (*10%*) of donations has no negative impact on male workers, while in the model (4), it significantly and negatively affects female participation (-31 percent, relative to the constant). For all the other levels of antisocial incentives, the negative effects (in relative terms) are quite similar across the two groups.

Result 2 (Labor supply and antisocial incentives) *Compared to treatment 0%, antisocial incentives attract a lower number of workers when 10%, 50%, 90%, or 100% of the bonus is donated.*

3.3 Intensive margin: Quantity

Building on the previous analysis, we now proceed with analyzing the first proxy for the worker's effort provision. Table A.2 assesses the impact of treatments on the *quantity* (Q) of keywords provided.

Model (1) shows how the number of keywords does not vary systematically across conditions. Under NRA *100%*, we detect a significant negative effect (-22 percent) compared to *0%*. This

[16] The size of the percent effect is computed as: $\frac{\beta_i}{\beta_0 Const.}$.

negative effect survives in Model (2), where we introduce the battery of controls. Altruism is marginally positively associated with the quantity, while the coefficient for females turns out positive, statistically significant, and economically meaningful (+20 percent).

To investigate potential differences in gender, Models (3) and (4) focus only on male and female workers, respectively. While male workers react negatively to UNICEF *100%* (Model 3), female workers show a milder negative reaction to the homologous NRA configuration (Model 4). The general mildly positive effect of altruism (Model 2) appears to be driven by female workers (model 4).

Result 3 (Quantity and social incentives) *Compared to treatment 0%, social incentives do not affect effort provision in terms of quantity when 10%, 50%, or 90% of the bonus is donated. Quantities are lower for antisocial incentives when 100% of the bonus is donated.*

3.4 Intensive margin: Completion time

While Table A.2 does not reveal any substantial effect on quantity in the different treatments, Table A.3 focuses on the *completion time* (T). This outcome measures how many seconds took the workers to complete the task.

It highlights some interesting differences. The average completion time in the baseline condition is 470 seconds (approximately 8 minutes, from starting the first to the last picture). All the different specifications show in a systematic way how workers exposed to UNICEF *10%*, *50%*, and *90%* are 210 seconds (approximately 3 minutes and 30 seconds) faster in completing their tasks compared to the workers in treatment *0%*.

This effect is robust even when we take into account the standard battery of control variables (Model 2), the number of keywords generated (Model 3 and 4), and when we control for the rating measure that captures the accuracy (Model 5, and 6) of the keywords provided (see below). This also holds in Models (7) and (8) when both factors are taken into account simultaneously.

Weak evidence, both in size and statistical significance, suggests a marginally faster performance in treatment NRA *100%*.

As far it concerns the control variables (Models 2, 4, 6, and 8), female workers seem, on average, to be somewhat quicker in completing the job compared to males. Analogously, it is possible to observe a positive association between altruism and speed in processing the task (a one standard deviation increase in altruism generates a positive effect of about 28 seconds). Older workers are somewhat slower than younger ones.

Result 4 (Speed and social incentives) *Compared to treatment 0%, prosocial incentives increase effort in terms of speed when 10%, 50%, or 90% of the bonus is donated. Speed remains largely unaffected for antisocial incentives.*

3.5 Intensive margin: Accuracy

Higher celerity in completing the keywords task might not be a desirable outcome if this means more sloppiness and errors in the output. For this reason, we also analyze the degree of *accuracy* (A) of the keywords by each worker.

Table A.4 shows how the average accuracy levels under the different UNICEF configurations, in which under *10%*, *50%*, and *90%* workers were significantly faster compared to *0%*, do not differ from the average level of accuracy applied in *0%* (Average rating: 2.5).

Under the NRA treatments, we observe negative coefficients that reveal a lower levels of accuracy for antisocial incentives. Their statistical significance tends to disappear when an individual's control variables are considered.

Coefficients for quantity (in Model 2) and accuracy (in Model 3) are both very small in their magnitudes and far from any conventional level of statistical significance. The higher speed in performing the task does not bring costs in terms of the accuracy of the output.

Result 5 (Accuracy and social incentives) *Compared to treatment 0%, prosocial incentives do not affect effort provision in terms of accuracy. Accuracy is lower for antisocial incentives when 50%, 90%, or 100% is donated.*

3.6 Intensive margin: Work efficiency

We conclude our analysis by analyzing the effect of the different treatments on the (standardized) multidimensional *work-efficiency* index (E). It condenses the complete set of outcomes we can observe in our experimental design in one single metric.

Model (1) in Table 4 shows how workers exposed to UNICEF 10%, 50%, and 90% generated a level of standardized productivity that is between .7 and .9 standard deviation higher than under the baseline condition.

No significant positive effects are detected under the NRA antisocial configurations. Results are confirmed when the set of control variables is taken into account in Model (2). Among the set of control variables, only the female dummy turns positive and statistically significant.

To shed more light on the gender effect, Models (3) and (4) consider males and females separately. The positive results registered under UNICEF *10%*, *50%*, and *90%* are confirmed for both the sub-samples, although the magnitudes of the effects in the female group are systematically and significantly higher.

Result 6 (Work efficiency and social incentives) *Compared to treatment 0%, prosocial incentives increase work efficiency when 10%, 50%, or 90% of the bonus is donated. Work efficiency remains largely unaffected for antisocial incentives.*

4 Discussion and conclusions

With a field experiment implemented in an online gig labor market, we study whether substituting a share of workers wage with a donation influences workers' labor supply and effort provision. We explore nonmonetary motivations of gig workers – who are an increasingly important part of labor markets.[17]

[17] A recent 2018 Deloitte report documented a 28-percent expected growth in gig-work volume in the next two years. Executives and practitioners are looking for ways to align corporate cultures and management practices with this emerging external workforce segment (Abbatiello et al., 2018; Butschek et al., 2019). It is urgent to examine how employer characteristics and practices influence the motivation of these non-traditional workers whose work experience is fundamentally different from the one of traditional in-house employees (Johnston and Land-Kazlauskas, 2018).

Table 4: Work efficiency (E) and social incentives

	Dependent variable: Work-efficiency index			
	(1)	(2)	(3)	(4)
Independent variables			-males-	-females-
UNICEF				
10%	0.841***	0.852***	0.259*	1.264***
	(0.143)	(0.139)	(0.140)	(0.205)
50%	0.892***	0.873***	0.590*	1.030***
	(0.191)	(0.187)	(0.304)	(0.238)
90%	0.705***	0.686***	0.487***	0.786***
	(0.158)	(0.158)	(0.180)	(0.200)
100%	0.002	-0.023	-0.150	0.078
	(0.047)	(0.055)	(0.099)	(0.066)
NRA				
10%	-0.040	-0.007	-0.069	-0.008
	(0.037)	(0.046)	(0.075)	(0.052)
50%	-0.057	-0.057	-0.037	-0.060
	(0.043)	(0.058)	(0.086)	(0.066)
90%	-0.060	-0.060	-0.087	-0.028
	(0.050)	(0.064)	(0.110)	(0.074)
100%	-0.081*	-0.115	-0.061	-0.187*
	(0.042)	(0.076)	(0.100)	(0.099)
Control variables				
Female		0.299***		
		(0.082)		
Age		-0.004	-0.008*	-0.001
		(0.004)	(0.004)	(0.005)
Altruism		-0.041	0.005	-0.058
		(0.039)	(0.045)	(0.056)
Risk: financial		0.053	-0.034	0.076
		(0.049)	(0.046)	(0.072)
Risk: ethical		0.036	0.007	0.085
		(0.044)	(0.059)	(0.067)
Machiavellianism		-0.010	-0.000	-0.014
		(0.040)	(0.046)	(0.056)
0%(Const.)	-0.324***	-0.374**	-0.004	-0.250
	(0.027)	(0.155)	(0.167)	(0.215)
N	497	497	171	326
R^2	0.181	0.204	0.152	0.236

Notes: Coefficients (robust standard errors in parentheses) of OLS model. Dependent variable: z-standardized work-efficiency index $E = z(Quantity \cdot \frac{Accuracy}{Completion\ Time})$. Altruism, Risk: financial, Risk: ethical, and Machiavellianism are z-standardized. Significance indicators: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In our field experiment, we investigate the workers' decisions to accept a specific contract and – if they accepted the job – their effort provision. The results of our field experiment show a non-negative effect of substituting the workers' wages with small and moderate prosocial donations on labor supply; once the prosocial donation becomes larger than the wage, labor supply drops. Workers selecting into the job perform more efficiently when parts of their wages are donated compared to a contract when no social incentive is involved. This increasing in efficiency is mainly due to workers increase their celerity of completing the task without making more mistakes. When social incentives do not align with the prosocial cause and involve an antisocial organization, labor supply drops when substituting even small shares of the wage with a donation. Workers selecting into the job for antisocial incentives do not show systematically different behavior compared to a contract when no social incentive is involved (if anything, they marginally lower the level of effort in terms of accuracy of the job).

The treatment effects of prosocial incentives on the labor supply explored in this paper highlight a different mechanism from those observed in the recent studies on prosocial incentives and CSR (e.g., Hedblom et al., 2019). Here, it has been shown that the mere information about the employer's CSR activities increases labor supply and work efficiency when the CSR activity is virtually costless for the worker. In this paper, we show that even when workers "have to put their effort where their mouth is" and face moderate prosocial incentives which reduce their wage, labor supply remains stable, but leads to an increase in work efficiency.

This paper provides evidence that it is not primarily the fact that workers create a positive externality (i.e., warm glow) with their efforts that drives the behavioral response to the prosocial incentive. In line with the laboratory experiment of Kajackaite and Sliwka (2017), who study the behavioral motives of agents working more for prosocial principals, and find that distributional concerns can be one driver of higher efforts of workers. In their experiment, workers (agents) are willing to increase their effort, but only when the principal is not wealthier. Under the assumption that the organization (here UNICEF) is perceived as the principal, a reduction in labor supply can be explained when donations (and reduction of wages, respectively) are too large.

Small or moderate prosocial incentives may have the advantage of attracting two types of worker: those who think their efforts are also meaningful (Cassar and Meier, 2018; Dur and van Lent, 2019) in helping the charitable cause, and those who are solely interested in their benefit and consequently drive up efforts at the intensive margin.

In addition to that, our experimental treatments on antisocial incentives document a "cold-shiver" effect on labor supply. For small and moderate donations to an antisocial cause, we observe a lower acceptance rate compared to the prosocial donations. Predicting all metrics separately and comparing them between prosocial and antisocial incentives yields a lower number of workers, quantity, velocity, accuracy, and work-efficiency for antisocial incentives (see Table 5).[18]

Of course, our study is not without limitations. The online labor market used in this paper is an ideal context in which to identify causal relationships of the selection of social incentive contracts. However, it is essential to note, that in these settings, workers are wholly anonymous

[18] A pairwise comparison for all treatments and metrics can be found in Table A.5 in the Appendix.

and separate from their employers. Future work could explore how less anonymous work relationships shape workers' attitudes towards social incentives. We also want to admit that we consider rather extreme cases of prosocial and antisocial incentives and beneficiaries of donations, either giving money from the worker to a particularly good cause (UNICEF) or bad cause (NRA). We also assume that the bonus is deducted directly from the workers' wages. These are both simplifications.

In many cases, organizations have multiple beneficiaries of corporate donations. Also, in most of the cases, companies donate to causes they think are good. Our experimental treatments involving antisocial incentives should be seen as a case where the values of workers and those of the respective CSR programs are at odds. Moreover, donations are not directly deducted from the workers' payment. Sometimes there are situations where corporate donations and workers' wages are in direct conflict, for example when firms promulgate profit-sharing, or any other form of performance-related pay. In these cases, the money given to social causes is directly deducted from the company's profits and cannot be spent on profit sharing or performance-related bonuses. These limitations somewhat restrict the generalizability of these findings to longer-term jobs with higher wages.

From a practical perspective, our findings suggest that moderate prosocial incentives can create a win-win situation for firms and society. Firms offering moderate prosocial incentives can lower their base wages and attract workers who work more efficiently and create positive externalities. Thus, prosocial incentives – allowing workers to donate small parts of their wages to prosocial causes – can be a way to align incentives with societal needs and boost performance at the same time.

Table 5: Comparing prosocial and antisocial incentives

	Dependent variable:				
Independent variables	(L)abor supply	(Q)uantity	Completion (T)ime	(A)ccuracy	Work (E)fficiency
1 if **NRA**	-0.182***	-5.365*	2.545***	-0.076**	-0.762***
	(0.031)	(2.955)	(0.430)	(0.035)	(0.085)
Constant (UNICEF)	0.336**	24.997	6.262***	3.098***	0.463
	(0.146)	(15.667)	(2.166)	(0.152)	(0.443)
Additional controls	Yes	Yes	Yes	Yes	Yes
N	833	409	409	409	409
R^2	0.183	0.050	0.101	0.183	0.178

Notes: Coefficients (robust standard errors in parentheses) of a linear probability model predicting all dimensions of labor supply (L) and effort provision (Quantity (Q), Completion Time (T), Accuracy (A), and Work-efficiency (E)). Control variables (omitted from display) are Altruism, Risk: financial, Risk: ethical, Machiavellianism, female, and indicators for the different bonus/donations constellations. Significance indicators: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

References

Abbatiello, A., Agarwal, D., and Bersin, J. (2018). Deloitte global human capital trends. 1.

America's Charities (2017). Snapshot 2017 - what U.S. employees think about workplace giving, volunteering, and CSR. Technical report.

Ariely, D., Bracha, A., and Meier, S. (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *American Economic Review*, 99(1):544–555.

Ashraf, N., Bandiera, O., and Lee, S. S. (2018). Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services. *Unpublished*.

Bagnoli, M. and Watts, S. G. (2017). Voluntary assurance of voluntary CSR disclosure. *Journal of Economics & Management Strategy*, 26(1):205–230.

Baron, D. P. (2007). Corporate social responsibility and social entrepreneurship. *Journal of Economics & Management Strategy*, 16(3):683–717.

Blais, A.-R. and Weber, E. U. (2006). A domain-specific risk-taking (DOSPERT) scale for adult populations. *Judgment and Decision Making*, 1(1):33–47.

Blanco, M., Engelmann, D., and Normann, H.-T. (2011). A within-subject analysis of preferences. *Games and Economic Behavior*, 72(06):321–338.

Burbano, V. (2016). Social responsibility messages and worker wage requirements: Field experimental evidence from online labor marketplaces. *Organization Science*, 27(4):1010–1028.

Burbano, V. (2019). Getting gig workers to do more by doing good: Field experimental evidence from online platform labor marketplaces. *Organization & Environment*, forthcoming.

Burbano, V. and Chiles, B. (2018). Corporate social responsibility and employee misconduct. *Working Paper*.

Butschek, S., González Amor, R., Kampkötter, P., and Sliwka, D. (2019). Paying gig workers - Evidence from a field experiment. *IZA Discussion Paper No. 12667*, (12667).

Calveras, A. and Ganuza, J.-J. (2016). The role of public information in corporate social responsibility. *Journal of Economics & Management Strategy*, 25(4):990–1017.

Carpenter, J. and Gong, E. (2016). Motivating agents: How much does the mission matter? *Journal of Labor Economics*, 34(1):211–236.

Cassar, L. and Meier, S. (2017). Intentions for doing good matter for doing well: The (negative) signaling value of prosocial incentives. *NBER Working Paper*, 24109(24109).

Cassar, L. and Meier, S. (2018). Nonmonetary incentives and the implications of work as a source of meaning. *Journal of Economic Perspectives*, 32(3):215–238.

Charities Aid Foundation (2018). Corporate giving by the FTSE 100 - bigger impact through better business. 1(January).

Charness, G., Cobo-Reyes, R., and Sánchez, Á. (2016). The effect of charitable giving on workers' performance: Experimental evidence. *Journal of Economic Behavior and Organization*, 131:61–74.

Charness, G., Gneezy, U., and Henderson, A. (2018). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, 149:74–87.

Charness, G. and Grieco, D. (2018). Creativity and incentives. *Journal of the European Economic Association*, 17(2):454–496.

Christie, R. and Geis, F. L. (1970). *Studies in Machiavellianism*. NY: Academic Press.

Conrads, J., Irlenbusch, B., Reggiani, T., Rilke, R. M., and Sliwka, D. (2016). How to hire helpers? Evidence from a field experiment. *Experimental Economics*, 19(3):577–594.

Dewatripont, M., Rochet, J.-C., and Tirole, J. (2015). *Balancing the banks: Global lessons from the financial crisis*. Number 9155 in Economics Books. Princeton University Press.

Dur, R. and van Lent, M. (2019). Socially useless jobs. *Industrial Relations*, 58(1):3–16.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

Financial Times (2011). Credit suisse to enforce charitable donations.

Förster, J., Higgins, E. T., and Bianco, A. T. (2003). Speed/accuracy decisions in task performance: Built-in trade-off or separate strategic concerns? *Organizational Behavior and Human Decision Processes*, 90(1):148–164.

Friedman, M. (1970). The social responsibility of firms is to increase its profits. *The New York Times*, pages 31–32.

Gall, T., Hu, X., and Vlassopoulos, M. (2019). Dynamic incentive effects of assignment mechanisms: Experimental evidence. *Journal of Economics & Management Strategy*, 28:687–712.

Gill, D. and Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, 102(1):469–503.

Gond, J.-P. and Nyberg, D. (2017). Materializing power to recover corporate social responsibility. *Organization Studies*, 38(8):1127–1148.

Goodman, J. K., Cryder, C. E., and Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of mechanical turk samples. *Journal of Behavioral Decision Making*, 26(3):213–224.

Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., and Bigham, J. P. (2018). A data-driven analysis of workers ' earnings on amazon mechanical turk. *Conference on Human Factors in Computing Systems*.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of Economic and Social Measurement, Volume 5, number 4*, pages 475–492. NBER.

Hedblom, D., Hickman, B. R., and List, J. (2019). Toward an understanding of corporate social responsibility: Theory and field experimental evidence. *NBER Working Paper*, (26222).

Imas, A. (2014). Working for the warm glow: On the benefits and limits of prosocial incentives. *Journal of Public Economics*, 114:14–18.

Johnston, H. and Land-Kazlauskas, C. (2018). Organizing on-demand representation, voice, and collective bargaining in the gig-economy. *Working Paper International Labour Organization*.

Kajackaite, A. (2015). If I close my eyes, nobody will get hurt: The effect of ignorance on performance in a real-effort experiment. *Journal of Economic Behavior & Organization*, 116:518–524.

Kajackaite, A. and Sliwka, D. (2017). Social responsibility and incentives in the lab: Why do agents exert more effort when principals donate? *Journal of Economic Behavior & Organization*, 142:482–493.

Kässi, O. and Lehdonvirta, V. (2018). Online labour index: Measuring the online gig economy for policy and research. *Technological Forecasting and Social Change*, 137:241–248.

Laske, K. and Schroeder, M. (2017). Quantity, quality and originality: The effects of incentives on creativity. *Working Paper of the German Economic Association*.

List, J. A. and Momeni, F. (2017). When corporate social responsibility backfires: Theory and evidence from a natural field experiment. *NBER Working Paper Series*, (24169).

Rushton, J. P., Chrisjohn, R. D., and Fekken, G. C. (1981). The altruistic personality and the self-report altruism scale. *Personality and Individual Differences*, 2(4):293–302.

Sauermann, J. (2016). Performance measures and worker productivity. *IZA World of Labor*, (5):260–271.

Schwartz, D., Keenan, E. A., Imas, A., and Gneezy, A. (2019). Opting-in to prosocial incentives. *Organizational Behavior and Human Decision Processes*.

Tonin, M. and Vlassopoulos, M. (2015). Corporate philanthropy and productivity: Evidence from an online real effort experiment. *Management Science*, 61(8):1795–1811.

Appendix

Theoretical considerations If we assume that a worker is comparing her part of the bonus with the part of the social cause then she might experience inequity aversion. Her utility can be represented as proposed by Fehr and Schmidt (1999):

$$U_i(B) = B_w - \alpha_w \cdot \max\{B_s - B_w\} - \beta_w \cdot \max\{B_w - B_s\}.$$

B_i with $i \in \{w, s\}$ represent the bonus given to the (w)orker or (s)ocial cause. It is assumed that $\alpha_w \geq \beta_w$ and $0 \leq \beta_w < 1$. The first assumption indicates that workers are experiencing a larger utility loss from disadvantageous than from advantageous inequity. Further non-negative values of β_w indicate that workers experience no utility gain from being better off than the social cause; while restricting $\beta_w \leq 1$ excludes implausible action to restore equity (burning money).

In Table A.1, we compute the range for α_w and β_w when workers should reject the job offer. Workers in *0%*and *10%* should only reject the job offer when $\beta_w \geq 1$ or 1.125, respectively. In *50%*, she would always accept. Workers in *90%*and *100%* will reject the job offer for values for $\alpha_w \geq 1/8$ or 0, respectively.

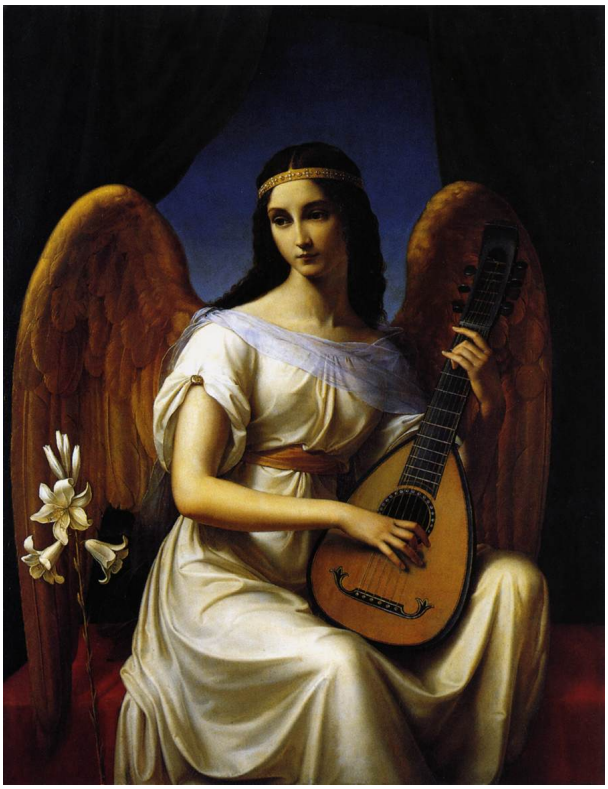
Consequently, we should observe an increasing number of workers rejecting the social incentives when disadvantageous inequality increases.

Table A.1: Inequity aversion and treatments

	Total Bonus B	Amount of B given to...		Worker rejects when ...	
		... worker	... social cause	... $\alpha_w \geq$	... $\beta_w \geq$
Treatment					
0%	\$ .50	\$ .50	\$ 0	-	1
10%	\$ .50	\$ .45	\$ .05	-	1.125
50%	\$ .50	\$ .25	\$ .25	-	-
90%	\$ .50	\$ .05	\$ .45	1/8	-
100%	\$ .50	0	\$ .50	0	-

Notes: Fehr and Schmidt (1999) and Blanco et al. (2011) estimate these parameters and show that the majority of participants in their subject pools have indeed non-negative values of α_i.

Figure A.1: Example picture "Mignon" (1828) by Friedrich Wilhelm von Schadow



Notes: This is one example picture that we used for the additional job. Gig workers have to describe ten comparable pictures by finding keywords. All pictures have been presented in a randomized order. One worker who provided six keywords on this picture stated: “Angel (Average accuracy by two independent raters: 4), Wings (4), Lillies (0), Music (3.5), Headband (3), Brunette (3).” A complete list of all pictures is available upon request. The keyword "Lillies" has received the lowest score since it is a spelling mistake and thus and indicator of inaccuarcy.

Table A.2: Quantity (Q) and social incentives

	Dependent variable: numbers of keywords			
Independent variables	(1)	(2)	(3) -males-	(4) -females-
UNICEF				
10%	1.562	2.613	-2.368	6.086
	(4.558)	(4.579)	(6.936)	(6.092)
50%	-0.417	-1.554	-10.083	2.717
	(4.600)	(4.540)	(6.204)	(5.947)
90%	1.113	-0.946	-2.108	0.654
	(6.738)	(6.770)	(12.864)	(8.067)
100%	4.167	1.863	-18.278***	6.947
	(6.724)	(6.505)	(6.704)	(8.000)
NRA				
10%	-3.084	-2.070	-2.741	-1.577
	(4.502)	(4.491)	(6.517)	(6.326)
50%	-2.123	-3.075	3.317	-4.255
	(4.577)	(4.654)	(6.156)	(6.197)
90%	-7.843	-9.486	-8.203	-8.976
	(6.097)	(5.872)	(8.995)	(7.533)
100%	-12.040**	-12.480**	-10.746	-15.394*
	(5.583)	(6.163)	(7.521)	(9.179)
Control variables				
Female		8.930***		
		(2.986)		
Age		0.108	-0.154	0.222
		(0.106)	(0.182)	(0.141)
Altruism		3.111*	-1.005	4.377**
		(1.632)	(2.484)	(2.103)
Risk: financial		-1.597	-2.486	-1.096
		(1.678)	(2.445)	(2.267)
Risk: ethical		1.094	0.691	2.828
		(1.876)	(2.375)	(2.665)
Machiavellianism		1.257	-0.652	1.560
		(1.491)	(2.696)	(1.838)
0%(Const.)	55.227***	45.490***	59.177***	48.673***
	(3.108)	(5.540)	(8.078)	(7.076)
N	497	497	171	326
R^2	0.013	0.042	0.052	0.051

Notes: Coefficients (robust standard errors in parentheses) of OLS model. Dependent variable: Number of keywords. Altruism, Risk: financial, Risk: ethical, and Machiavellism are z-standardized.
Significance indicators: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.3: Completion time (T) and social incentives

	Dependent variable: Completion time (in seconds)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Independent variables								
UNICEF								
10%	-214.586***	-219.016***	-219.551***	-227.734***	-214.244***	-218.946***	-219.189***	-227.660***
	(41.417)	(40.201)	(40.191)	(38.707)	(41.425)	(40.226)	(40.194)	(38.710)
50%	-197.600***	-197.432***	-196.274***	-192.246***	-197.371***	-197.383***	-196.031***	-192.194***
	(42.717)	(41.132)	(39.244)	(37.698)	(42.748)	(41.154)	(39.245)	(37.702)
90%	-218.815***	-223.419***	-222.351***	-220.261***	-218.367***	-223.293***	-221.877***	-220.128***
	(44.964)	(44.912)	(40.214)	(40.227)	(44.967)	(44.970)	(40.052)	(40.189)
100%	26.493	29.297	13.247	23.079	27.172	29.597	13.965	23.397
	(72.654)	(71.129)	(71.049)	(68.569)	(72.721)	(71.327)	(71.154)	(68.843)
NRA								
10%	-19.748	-27.578	-9.945	-20.670	-18.495	-27.194	-8.614	-20.263
	(46.096)	(45.411)	(40.819)	(39.596)	(46.182)	(45.491)	(40.746)	(39.508)
50%	8.170	12.621	14.919	22.880	10.143	13.150	17.012	23.441
	(46.218)	(44.457)	(45.617)	(43.626)	(46.502)	(44.765)	(45.629)	(43.653)
90%	-64.801	-68.821	-39.873	-37.169	-62.275	-68.296	-37.191	-36.613
	(48.090)	(46.378)	(42.039)	(40.578)	(47.704)	(46.031)	(41.664)	(40.329)
100%	-119.017***	-112.475**	-80.750**	-70.833*	-116.432***	-112.251**	-78.002**	-70.594*
	(44.339)	(44.438)	(38.611)	(41.472)	(44.616)	(44.448)	(38.377)	(41.335)
Control variables								
Female		-22.642		-52.440*		-22.687		-52.488*
		(28.853)		(27.188)		(28.965)		(27.289)
Age		4.218***		3.857***		4.226***		3.866***
		(1.253)		(1.161)		(1.253)		(1.161)
Altruism		-28.469**		-38.850***		-28.322**		-38.695***
		(12.949)		(12.035)		(13.119)		(12.109)
Risk: financial		23.745		29.075*		24.009		29.355*
		(16.534)		(15.385)		(16.425)		(15.330)
Risk: ethical		-17.343		-20.992		-16.896		-20.519
		(15.226)		(14.097)		(16.224)		(14.933)
Machiavellianism		-5.270		-9.464		-5.247		-9.441
		(14.141)		(12.933)		(14.142)		(12.939)
Quantity (Q)			3.178***	3.337***			3.179***	3.337***
			(0.493)	(0.488)			(0.490)	(0.488)
Accuracy(A)					12.046	4.437	12.775	4.700
					(31.494)	(38.502)	(25.881)	(31.896)
0%(Const.)	469.955***	329.176***	294.419***	177.385***	439.370***	317.715***	261.956***	165.241*
	(30.924)	(60.090)	(38.406)	(62.695)	(84.688)	(110.349)	(69.190)	(92.743)
N	497	497	497	497	497	497	497	497
R^2	0.124	0.166	0.239	0.290	0.124	0.166	0.239	0.290

Notes: Coefficients (robust standard errors in parentheses) of OLS model. Dependent variable: seconds of work spent to process the task. Altruism, Risk: financial, Risk: ethical, and Machiavellianism are z-standardized. Significance indicators: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.4: Accuracy (A) and social incentives

	Dependent variable: Accuracy rating			
	(1)	(2)	(3)	(4)
Independent variables				
UNICEF				
10%	-0.028	-0.028	-0.023	-0.013
	(0.058)	(0.057)	(0.062)	(0.059)
50%	-0.019	-0.019	-0.014	-0.009
	(0.061)	(0.061)	(0.065)	(0.060)
90%	-0.037	-0.037	-0.031	-0.026
	(0.062)	(0.062)	(0.067)	(0.062)
100%	-0.056	-0.056	-0.057	-0.068
	(0.092)	(0.093)	(0.092)	(0.095)
NRA				
10%	-0.104	-0.104	-0.104	-0.086
	(0.063)	(0.064)	(0.064)	(0.059)
50%	-0.164**	-0.164**	-0.164**	-0.120*
	(0.078)	(0.079)	(0.078)	(0.069)
90%	-0.210**	-0.210**	-0.208**	-0.118
	(0.104)	(0.103)	(0.103)	(0.087)
100%	-0.215**	-0.215**	-0.211**	-0.050
	(0.090)	(0.092)	(0.092)	(0.100)
Control variables				
Quantity (Q)		-0.001		-0.001
		(0.001)		(0.001)
Completion time (T)			0.001	0.001
			(0.001)	(0.001)
Female				0.011
				(0.039)
Age				-0.002
				(0.002)
Altruism				-0.033*
				(0.019)
Risk: financial				-0.060***
				(0.020)
Risk: ethical				-0.100***
				(0.022)
Machiavellianism				-0.005
				(0.016)
0%(Const.)	2.539***	2.541***	2.527***	2.582***
	(0.046)	(0.071)	(0.064)	(0.088)
N	497	497	497	497
R^2	0.030	0.030	0.031	0.176

Notes: Coefficients (robust standard errors in parentheses) of OLS model. Dependent variable: accuracy rating on a 5-likert scale (0=not accurate; 4=very accurate). Altruism, Risk: financial, Risk: ethical, and Machiavellianism are z-standardized. Significance indicators: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.5: Overview of average values and pairwise comparisons

					NRA		
a) Labor supply (L)			*0%*	*10%*	*50%*	*90%*	*100%*
		(frac.)	.79	.59	.48	.38	.16
	0%	.79	-	**.001**	**.000**	**.000**	**.000**
UNICEF	*10%*	.75	.484	**.012**	.117	**.002**	**.000**
	50%	.77	.649	**.808**	**.000**	.145	**.000**
	90%	.44	**.000**	**.000**	**.000**	.341	**.001**
	100%	.35	**.000**	**.000**	**.000**	.169	**.002**
b) Quantity (Q)			*0%*	*10%*	*50%*	*90%*	*100%*
		(mean)	55.23	52.14	53.10	47.38	43.19
	0%	55.23	-	.2709	.8073	**.0394**	.1213
UNICEF	*10%*	56.79	.7403	.3023	.3854	.3006	.2873
	50%	54.81	.6712	.5014	.6619	.0804	.1366
	90%	56.43	.3152	.3427	.6567	.3680	.9233
	100%	59.39	.7080	.9154	.4874	.3311	.1498
c) Completion time (T)			*0%*	*10%*	*50%*	*90%*	*100%*
		(mean)	469.8	450	478.2	405	351
	0%	469.8	-	.6852	.4409	.3331	.1304
UNICEF	*10%*	255.6	**.0000**	**.0000**	.2586	.5658	.2479
	50%	272.4	**.0000**	.7339	**.0000**	.0907	**.0372**
	90%	251.4	**.0000**	.8962	.6463	**.0002**	.5538
	100%	496.2	.7599	**.0000**	**.0000**	**.0000**	.2986
d) Accuracy (A)			*0%*	*10%*	*50%*	*90%*	*100%*
		(mean)	2.54	2.44	2.38	2.33	2.32
	0%	2.54	-	**.0336**	**.0182**	**.0421**	**.0087**
UNICEF	*10%*	2.51	.3174	.1468	.4961	.7826	.0981
	50%	2.52	.5303	.7098	**.0426**	.9645	.3847
	90%	2.50	.3068	.9307	.6524	.2492	.4349
	100%	2.48	.7100	.7615	.9423	.6402	**.0223**
e) Work efficiency (E)			*0%*	*10%*	*50%*	*90%*	*100%*
		(mean)	-.33	-.36	-.38	-.38	-.39
	0%	-.33	-	.5610	.2627	.1708	.5230
UNICEF	*10%*	.50	**.0000**	**.0000**	.6163	.3949	.8610
	50%	.56	**.0000**	.6060	**.0000**	.7437	.8120
	90%	.39	**.0000**	.9980	.5779	**.0000**	.5631
	100%	.32	.6470	**.0000**	**.0001**	**.0000**	.2792

Notes: On the diagonals we display p-values of comparisons between UNICEF and NRA. Below (above) the diagonal comparisons are made between different bonus/donation compositions with the same beneficiary (either UNICEF or NRA). For comparisons of a) Labor supply we use a χ^2 - test. b) Quantity, c) Completion time (in seconds), d) Accuracy (rating scale), and e) Work efficiency (z-standardized index) are tested with a MWU-test. p-values are two-sided.