

Discussion Papers of the  
Max Planck Institute for  
Research on Collective Goods  
2020/10



**Monetary and Social Incentives  
in Multi-Tasking: The Ranking  
Substitution Effect**

**Matthias Stefan  
Jürgen Huber  
Michael Kirchler  
Matthias Sutter  
Markus Walzl**

**MAX PLANCK**  
SOCIETY





# **Monetary and Social Incentives in Multi-Tasking: The Ranking Substitution Effect**

**Matthias Stefan / Jürgen Huber / Michael Kirchler / Matthias Sutter / Markus Walzl**

May 2020

# Monetary and Social Incentives in Multi-Tasking: The Ranking Substitution Effect\*

Matthias Stefan, Jürgen Huber, Michael Kirchler,  
Matthias Sutter, and Markus Walzl<sup>†</sup>

May 5, 2020

## Abstract

Rankings are prevalent information and incentive tools in labor markets with strong competition for talent. In a dynamic model of multi-tasking and an accompanying experiment with financial professionals, we identify hidden ranking costs when performance in one task is incentivized and ranked while another prosocial task is not: (i) a ranking influences behavior if individuals lag behind: they spend more total effort and substitute effort in the prosocial task with effort in the ranked task; (ii) those ahead in the ranking spend less total effort and lower relative effort in the ranked task. Implications for incentive schemes are discussed.

JEL: C93, D02, D91

Keywords: multi-tasking decision problem, rank incentives, framed field experiment, finance professionals.

---

\*Our experimental software, data files, and the analysis script file are stored on an OSF repository and will be made publically available once the paper has been accepted for publication.

We thank Nora Szech, Christian König, and seminar and conference participants at the SFB “Credence Goods, Incentives, and Behavior”-Workshop 2019 Innsbruck and the ZEW Workshop on Market Design 2019 for very valuable comments on earlier versions of this paper. Financial support from the Austrian Science Fund (FWF START-grant Y617-G11, SFB F63) and the Swedish Research Council (grant 2015-01713) is gratefully acknowledged. This study was ethically approved by the IRB of the University of Innsbruck.

<sup>†</sup>Stefan: University of Innsbruck, Department of Banking and Finance, Universitätsstrasse 15. E-mail: matthias.stefan@uibk.ac.at

Huber: University of Innsbruck, Department of Banking and Finance, Universitätsstrasse 15. E-mail: juergen.huber@uibk.ac.at

Kirchler: University of Innsbruck, Department of Banking and Finance, Universitätsstrasse 15, 6020 Innsbruck. E-mail: michael.kirchler@uibk.ac.at.

Sutter: Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Strasse 10, Bonn, and University of Cologne, Department of Economics. E-mail: matthias.sutter@coll.mpg.de.

Walzl: University of Innsbruck, Department of Economics, Universitätsstrasse 15. E-mail: markus.walzl@uibk.ac.at.

# 1 Introduction

Monetary incentives and (non-pecuniary) social comparisons like rank incentives are among the most prevalent incentive structures in companies and public institutions (Coles et al., 2018). Examples include combinations of fixed salaries and bonus payments, relative-performance evaluations (tournaments), but also institutional designs that make use of social (peer) comparison to promote employee performance. However, numerous incidences of moral hazard, outright fraud, and bad performance indicate that ill-designed incentive structures can lead to antisocial behavior and negative externalities for the company and society as a whole. Surprisingly, little scientific evidence exists on trade-off decisions between individual monetary and rank incentives on the one hand and promoting prosocial behavior on the other hand.

In this paper, we narrow this research gap and investigate how monetary incentives and rank incentives (i.e., rankings) influence effort in a multi-tasking environment with a trade-off between one’s own payment and a prosocial activity. Ranking agents’ success in the task that determines her payment could serve as an additional incentive (if a good rank is regarded as desirable) or disincentive (if a good rank suggests less concern for the prosocial activity). In our experiment with financial professionals we find a *ranking substitution effect*: Those professionals that are ahead in the ranking reduce total effort and substitute relative effort spent for their own payment by spending more effort in the prosocial activity – i.e., the impact of a ranking is similar to a reduced piece-rate. In contrast, those lagging behind increase total effort and substitute by spending more effort for their own payment and less for the prosocial activity – i.e., the ranking works like an additional piece-rate. Hence, the benefits of a ranking come at a cost: While some individuals are motivated to spend more total effort, others are disincentivized, and while some act more prosocially, others focus more on the payoff-relevant task.

Our study is motivated by real-world anecdotal evidence on the role of monetary incentives and social comparison. To name just two examples: First, there has been a widespread debate on the “Wall Street culture” of incentive schemes in the finance industry. While these incentive schemes are pecuniary by nature, they also include a strong element of social comparison of salary among peers. In some sub-sectors like the (hedge) fund industry, professionals’ salaries are a convex function of past performance relative to other fund managers, resulting in a tournament incentive structure (Brown et al., 1996; Sirri and Tufano, 1998; Kaniel and Parham, 2017). These incentive schemes, however, have been identified as one of the reasons for the financial crisis in 2008 and of excessive risk taking in the finance industry in general (Rajan, 2006; Diamond and Rajan, 2009; Kirchler et al., 2018). This debate reflects public worries about the potential detrimental effects of monetary and rank incentives “gone wild” by generating

negative externalities. It also shows that prosocial behavior, such as contributing to a public good like financial stability, can be potentially overruled by the individual aspiration for high social status and monetary payments. Especially the financial crisis of 2008/09 has demonstrated glaring examples of excessive risk-taking, and of the public being harmed, by having to bail out financial institutions ruined by their well-paid managers.<sup>1</sup> Similar to the set-up we study, finance professionals make trade-off decisions between a high monetary payment (and thereby outsmart peers) and contributing to the public good of financial stability.

Second, publication merits in academia as a form of social comparison are not directly related with researchers' monetary incentives. While it is true that the scientific reputation of some positions and their payment are positively correlated, the hunt for top-ranked publications can hardly be explained solely by monetary incentives, in particular not for tenured senior researchers. As in the finance industry, it has been argued that status seeking may crowd-out intrinsic research motivation (Osterloh and Frey, 2015) and may be one of the reasons for misconduct and sabotage among researchers (Anderson et al., 2007; Fanelli, 2010). Similar to the set-up we examine, researchers have to trade off between high publication reputation and contributing to the public good of service to the community and students. We tackle the question how such trade-off decisions are made under different incentives-regimes and what the expected social impact is.

Both examples demonstrate the motivation behind our study: First, monetary incentives and social comparison in real life are highly relevant and can become the main driving force of employees. Second, social comparison in a multi-tasking environment can complement or substitute monetary incentives and motivations for prosocial activity depending on how the ranking is interpreted. As a consequence, rankings may undermine or promote prosocial behavior and thereby create positive and negative externalities to institutions and the society (Bénabou and Tirole, 2006).

In this paper, we introduce a dynamic multi-tasking model of an agent who derives utility from monetary payments in one task, but also from a prosocial task, thus facing a trade-off decision.<sup>2</sup> Over time, the agent can highlight or balance potentially conflicting motives as suggested by the literature on social psychology (see e.g., Fishbach et al., 2009). As a variation we introduce a ranking on the monetarily incentivized task (payment), which informs the agent about her performance in this task relative to other agents – but implicitly also about the intensity of her prosocial activity. Depending on how the agent interprets the ranking, it works as an additional incentive or disincentive for the monetarily incentivized task: If a good rank is desirable, it increases total effort and the fraction of effort spent for the ranked task. If a good

---

<sup>1</sup>One example is Richard Fuld, former CEO of Lehman Brothers, who received USD 71.9 million in bonus and salary in 2008, the year the bank went bankrupt. Over the five years before the collapse, Fuld earned USD 354 million.

<sup>2</sup>Similar to Bénabou and Tirole (2016), one could interpret the first task as observable to the employer or the company.

rank is costly because it is regarded as an signal (to the self or others) of low prosocial activity, it lowers total effort and the fraction of effort spent into the ranked task. If the agent balances the conflicting motives over time, we expect a ranking to be a disincentive for the ranked task if the agent’s performance was ranked highly in previous periods and an incentive for the ranked task in case of a poor ranking in previous periods.

We tested the hypotheses in a controlled online experiment with 286 internationally operating finance professionals from high-skilled areas (e.g., fund management, private banking, trading). They represent a suitable subject pool as they are exposed to various forms of social comparison in their profession. The professionals had to solve items of an IQ-related test – i.e., the Raven’s Advanced Progressive Matrices test (Raven, 2000) – and had to choose whether to solve these items for their own payment (selfish activity) or for donations for measles vaccine to UNICEF (the prosocial activity). In a 2x3 factorial experimental design, we varied the private display of a peer-ranking (displayed or not) and the level of monetary incentives (piece-rates) – where both incentives are only relevant for the items solved for oneself and not for the items solved for donation. Moreover, we also asked for participants’ prosocial and competitive preferences to differentiate between types of agents.

As our main contribution, we show that the introduction of a ranking on the monetarily incentivized activity leads to a “ranking substitution effect”: Those professionals that are ahead in the ranking substantially reduce total effort. They also substitute relative effort spent for their own payment by putting more effort into the prosocial activity. In contrast, those lagging behind increase total effort and additionally substitute by spending more effort for their own payment and less for the prosocial activity. Finally, we find evidence for another substitution effect, as higher monetary incentives lead to less effort provided for the prosocial activity and more for the own payment. We discuss our results in light of the optimal design of incentive schemes and labor market contract regulations to overcome moral hazard and adverse selection in (imperfectly competitive) labor markets for managerial talent (Bénabou and Tirole, 2016).

Our paper particularly contributes to literature on social comparison (Festinger, 1954; Bandiera et al., 2010; Cohn et al., 2015) and status (Moldovanu et al., 2007). Various studies disentangling rank incentives (social comparison) from monetary incentives show the effect of non-incentivized rankings on performance (Tran and Zeckhauser, 2012; Barankay, 2015), portfolio choice (Dijk et al., 2014), risk taking (Kuziemko et al., 2014; Kirchler et al., 2018), and market prices (Ball et al., 2001). With regards to effort provision, the literature reports varying effects of rankings, ranging from an overall increase in effort (Azmat and Iriberry, 2010; Blanes-i-Vidal and Nossol, 2011; Tran and Zeckhauser, 2012) to diverse effects depending on, for instance, expectations, current rank, and principal agent relationships (Al-Ubaydli and List, 2015; Kuhnen and Tymula, 2012; Gill et al., 2019).<sup>3</sup> Inquiries into optimal contracts also relate

---

<sup>3</sup>Moreover, peer effects need not necessarily be part of an explicit incentive scheme or an explicitly designed ranking, but can also emerge rather naturally (Mas and Moretti, 2009).

to insights into incentive schemes (e.g., Baker, 1992; Bénabou and Tirole, 2016). Our inquiry is also related to studies on self-image concerns (Bénabou and Tirole, 2006), which, by their very nature, are often not straightforward (Ariely et al., 2009; Falk and Szech, 2019). The trade-off choice in our model can also be interpreted as a balancing of the desire for a positive self-image due to observed rank and their desire for a positive self-image that stems from contributing to a prosocial activity. This also relates to the literature discussing how social comparison and monetary incentives can lower prosocial behavior or even promote misconduct (Shleifer, 2004; Charness et al., 2014).

Importantly, while most existing studies mainly focus on effort in one domain, we extend the literature by studying a multi-tasking problem where subjects can distribute their effort between a selfish and a prosocial activity. The novel aspect of our design is that subjects can distribute effort between two activities (i.e., a feature of most real-world decisions) allowing us to analyze implications for total effort and for substitution effects between both activities.<sup>4</sup> Moreover, our dynamic model adds to the literature on highlighting and balancing of (conflicting) motives over time in social psychology (see, e.g., Fishbach et al., 2009).

## 2 The Model

A risk neutral agent decides how much effort to spend on two different domains (effort  $a \geq 0$  into Domain  $A$  and effort  $b \geq 0$  into Domain  $B$ ) in each of  $t = 1, \dots, T$  periods. In a given period  $t$ , efforts  $a_t$  and  $b_t$  generate convex costs  $(a_t + b_t)^2/2$ .<sup>5</sup> The activity in Domain  $A$  is not verifiable but agents may be intrinsically motivated to spend effort  $a$  – in our setting, because effort  $a$  generates a positive externality. Output in Domain  $B$  is verifiable and agents spend effort in response to linear contracts with fixed payment  $z$  and piece-rate  $y$ . Additionally, the agent may receive feedback on her effort in Domain  $B$  in comparison to a reference group. To be specific, assume that, with effort  $b$ , the agent is better than  $R(b)$  other agents.  $R(b)$  is weakly monotone increasing in  $b$ , i.e., the higher  $b$  the better is the agent’s rank relative to a given reference group of agents. In sum, three motivational factors may therefore shape the agent’s behavior: (1) the positive externality (in our experiment a donation to UNICEF), (2) the monetary payment induced by the linear contract, and (3) the social comparison as introduced by the ranking of effort in Domain  $B$ .<sup>6</sup> As in Fishbach et al. (2009) we allow the agent to balance (or highlight) these motives over time: Denote the individual’s choice of  $a$  and  $b$  in period  $t$  by  $a_t$  and  $b_t$ ,

---

<sup>4</sup>For instance, literature on crowding-out of motivation by extrinsic incentives, too, mainly focuses on effort in one domain where subjects are (initially) intrinsically motivated for (see e.g., Gneezy and Rustichini, 2000a,b; Mellström and Johannesson, 2008; Bowles and Polania-Reyes, 2012).

<sup>5</sup>Our findings would remain unaltered for a more general convex costs functions as long as efforts  $a_t$  and  $b_t$  are substitutes. Given the time constraint in our experiment outlined below, it seems natural to assume convexity - but as long as the returns to efforts  $a_t$  and  $b_t$  are strictly concave this assumption is dispensable.

<sup>6</sup>In an anonymized setting, (1) and (2) are straightforward implications of standard other-regarding preferences (see, e.g., Fehr and Schmidt (1999) and Bolton and Ockenfels (2000)) while (3) could be the result of self-image concerns (see, e.g., Bénabou and Tirole (2006)).

and the individual's effort path until period  $t$  by  $a(t) = a_1, \dots, a_t$  and  $b(t) = b_1, \dots, b_t$ . The individual's (aggregate) utility at period  $t$  from the choice of  $(a(t), b(t))$  in periods  $\tau = 1, \dots, t$  with an (expected per-period) bonus  $y$  and an (expected per-period) fixed payment  $z$  is<sup>7</sup>

$$U_t(a(t), b(t); y, z) = \alpha\Phi\left(\sum_{\tau=1 \dots t} a_\tau\right) + \beta\Psi\left(\sum_{\tau=1 \dots t} R(b_\tau)\right) + b_t y + z - \sum_{\tau=1 \dots t} \frac{(a_\tau + b_\tau)^2}{2}.$$

$\alpha$  measures how relevant the individual considers the donation.  $\Phi$  is an increasing, weakly-concave and twice differentiable function of aggregate effort spent in Domain A. If  $\Phi$  is linear, efforts into Domain A in periods before  $t$  have no impact on the marginal benefit of effort into Domain A in period  $t$ . If  $\Phi$  is (strictly) concave, marginal benefits in period  $t$  (strictly) decrease in aggregate efforts spent in previous periods (e.g., due to moral licensing or other reasons for balancing of various motives).  $\beta$  measures the strength of the individual's competitive concerns, i.e.,  $\beta > 0$  if the individual is able to compare her performance to other agents' performance in the presence of a ranking and actually cares for such a social comparison.  $\Psi$  is a weakly-concave and twice differentiable function of  $R(b_t)$ . If  $\Psi$  is linear, efforts into Domain B in periods before  $t$  have no impact of the marginal benefit of effort into Domain B in period  $t$ . If  $\Psi$  is concave, marginal benefits in period  $t$  decrease in aggregate efforts spent in previous periods (e.g., the more effort spent in previous periods and - as a consequence - the better the rank in the previous periods, the lower is the incentive to spend effort in Domain B in period  $t$ ). In particular,  $\Psi$  may be non-monotone in  $R(b)$ . E.g., if  $\Psi$  attains a global maximum at  $R^*$ , the social comparison introduces an additional marginal benefit of  $b$  (similar to a higher piece-rate) for  $R(b) < R^*$  and an additional marginal cost of  $b$  (similar to a lower piece-rate) for  $R(b) > R^*$ .

**Relevance of donation:** The stronger the individual's intrinsic motivation to spend effort  $a$  (e.g., the more relevant the positive externality generated by  $a$  is to the individual), the higher is  $\alpha$ , and the higher is the marginal benefit of  $a$  which increases the optimal  $a_t^*$  and (as  $a$  and  $b$  are substitutes) decreases  $b_t^*$  in every period  $t$ . As the direct effect of  $\alpha$  on  $a_t$  is stronger than the indirect (substitution) effect of  $\alpha$  on  $b_t$ , the sum of the two efforts increases (as long as  $\Psi$  is strictly concave) while the fraction of efforts spent on Domain B decreases. (all proofs can be found in Appendix A).

**Proposition 1** For all  $t \in \{1, \dots, T\}$ , (i)  $\frac{da_t^*}{d\alpha} > 0$ , (ii)  $\frac{db_t^*}{d\alpha} < 0$ , (iii)  $\frac{d(a_t^* + b_t^*)}{d\alpha} > 0$  in the presence of a ranking and if  $\Psi$  is strictly concave and 0 in the absence of a ranking and if  $\Psi$  is linear, and (iv)  $\frac{db_t^*/(a_t^* + b_t^*)}{d\alpha} < 0$ .

<sup>7</sup>For instance, in our experiment, a participant is paid for choices made in one (out of four) periods which is randomly determined. Thus, in a given period, the participant expects to get one fourth of the piece-rate and one fourth of the fixed payment.

**Impact of piece-rates:** Increasing the piece-rate  $y$  yields a higher marginal benefit of  $b$  which increases  $b_t^*$  and (as  $a$  and  $b$  are substitutes) decreases  $a_t^*$ . As the direct effect of  $y$  on  $b$  is stronger than the indirect (substitution) effect of  $y$  on  $a$ , the sum of the two efforts increases (as long as  $\Phi$  is strictly concave) while the fraction of efforts spent on Domain  $B$  also increases.

**Proposition 2** For all  $t \in \{1, \dots, T\}$ , (i)  $\frac{db_t^*}{dy} > 0$ , (ii)  $\frac{da_t^*}{dy} < 0$ , (iii)  $\frac{d(a_t^*+b_t^*)}{dy} > 0$  if  $\Phi$  is strictly concave and 0 if  $\Phi$  is linear, and (iv)  $\frac{db_t^*/(a_t^*+b_t^*)}{dy} > 0$ .

Observe that whenever  $\Phi$  ( $\Psi$ ) is strictly concave, marginal returns of  $a$  ( $b$ ) decrease in  $t$ . As marginal costs remain unaltered across periods, this typically induces diminishing efforts.

**Impact of a ranking:** Denote the unique global maximum of  $\Psi$  on  $[0, \infty)$  by  $R^*$ . As for the extreme cases,  $R^* = 0$  implies that any rank but the last introduces a cost for the agent. If  $\Psi$  is monotone increasing (and  $R^* = \infty$ ), the presence of the ranking unambiguously motivates the agent to spend more effort in Domain  $B$  – the impact of the ranking is identical to the impact of a higher piece-rate.

**Corollary 1** Suppose  $\Psi$  is monotone increasing. Then, (i)  $\frac{db_t^*}{d\beta} > 0$ , (ii)  $\frac{da_t^*}{d\beta} < 0$ , (iii)  $\frac{d(a_t^*+b_t^*)}{d\beta} > 0$  if  $\Phi$  is strictly concave and 0 if  $\Phi$  is linear, and (iv)  $\frac{db_t^*/(a_t^*+b_t^*)}{d\beta} > 0$ .

If, however,  $\Psi$  is non-monotone (i.e.,  $0 < R^* < \infty$ ), the presence of a ranking in period  $t$  resembles an additional incentive to spend effort in Domain  $B$  for  $\sum_{\tau=1\dots t} R(b_\tau) < R^*$  (i.e., if a good rank is desirable because only bad ranks have been achieved so far) and an additional cost of effort in Domain  $B$  for  $\sum_{\tau=1\dots t} R(b_\tau) > R^*$  (i.e., if a good rank is not desirable any more since good ranks have been achieved in the past). As a consequence, the impact of the ranking is similar to a higher piece-rate in the former case and to a lower piece-rate in the latter case.

**Proposition 3** Let  $t \in \{1, \dots, T\}$ . For  $\sum_{\tau=1\dots t} R(b_\tau) < R^*$ , (i)  $\frac{db_t^*}{d\beta} > 0$ , (ii)  $\frac{da_t^*}{d\beta} < 0$ , (iii)  $\frac{d(a_t^*+b_t^*)}{d\beta} > 0$  if  $\Phi$  is strictly concave and 0 if  $\Phi$  is linear, and (iv)  $\frac{db_t^*/(a_t^*+b_t^*)}{d\beta} > 0$ ; for  $\sum_{\tau=1\dots t} R(b_\tau) > R^*$ , (i)  $\frac{db_t^*}{d\beta} < 0$ , (ii)  $\frac{da_t^*}{d\beta} > 0$ , (iii)  $\frac{d(a_t^*+b_t^*)}{d\beta} < 0$  if  $\Phi$  is strictly concave and 0 if  $\Phi$  is linear, and (iv)  $\frac{db_t^*/(a_t^*+b_t^*)}{d\beta} < 0$ ;

## 3 The experiment

### 3.1 Experimental design and treatments

In our online experiment, subjects had to solve items of the Raven’s Advanced Progressive Matrices test (APM; Raven, 2000). In each of the APM items, subjects had to recognize the geometric pattern in an unfinished diagrammatic puzzle and identify the missing element. The main objective of the APM is to measure subjects’ ability to solve novel problems, which is why it is also used as a measure of IQ (see Figure 1 for an example of an APM item and the full experimental instructions in Section C in the Appendix). One advantage of APM is that subjects are expected to be intrinsically motivated to obtain a higher rank in an intelligence test (Falk and Szech, 2019). Moreover, recent evidence shows that performance in IQ tests and related tasks are not only a measure of ability but also effort, and, thus, do indeed respond to incentives (Borghans et al., 2013; Gneezy et al., 2019).

In particular, subjects participated in four periods of two minutes each, consisting of two treatments (see details below). The order of items was randomized as follows: First, we randomized the APM items into two sequences which were then used for all participants. Then, for each subject a random draw decided which sequence occurred in which block (which is then randomly assigned to one within-subject treatment, i.e., with and without ranking; see below). In order to achieve a fair comparison, we kept the same two random sequences for all subjects and only randomly varied their ordering between blocks with and without ranking. In our experiment, participants could provide effort in two domains. Domain A (i.e., the prosocial activity) was not incentivized for the subject, but associated with a positive externality that is expected to generate an intrinsic prosocial motivation. In particular, we donated €10 to UNICEF for measles vaccination for each solved item for *Domain A*, which was described in detail in the instructions and was therefore public knowledge (see Figure 1 for a screenshot of the instructions outlining the prosocial activity; for a similar design see also Kirchler et al., 2016). For *Domain B*, in contrast, each subject received a certain piece-rate ( $y$ ) for each solved item for individual payment, which varied across treatments (€5, €10, or €15). Subjects had to decide for each APM item whether to solve it for Domain A or B *before* it was shown. Due to the random sequence, subjects could not infer the difficulty of the next item to be solved.

We set up a 2x3 factorial treatment design. Because of the large number of treatments for a framed field experiment (Harrison and List, 2004), we implemented a within-between-subjects design (see Table 1 for details).

As first treatment variation we modified whether a ranking was shown to the participants or not. We implemented this treatment variation as a within-subject design, i.e., each subject participated in a block of two periods with and in a block of two periods without a ranking in randomized order. The ranking was displayed at the end of each of the two periods within the block of treatments with ranking, based on a subject’s total number of correctly solved Raven

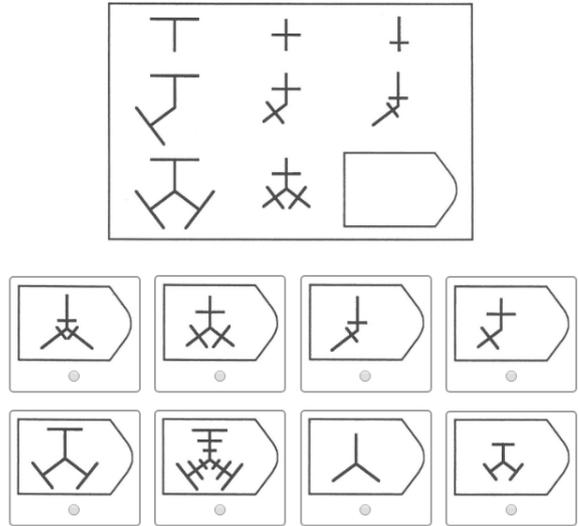
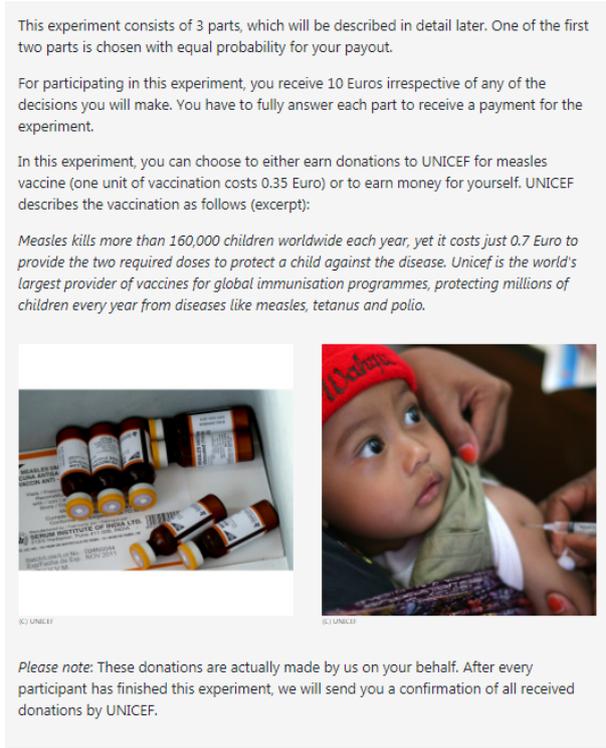


Figure 1: Left panel: Screenshot of the instruction screen outlining the explanation of the donation to measles vaccination to UNICEF (i.e., the prosocial activity). Right panel: Example of one APM item.

matrices for Domain B within two minutes. In particular, we showed the rank among 13 peers (12 plus the respective subject), but no further details about their performance. We pre-sampled the 12 professionals from the same subject pool in a first wave to constitute the ranking. They were invited to participate in an experiment that was identical to our main experiment except for two variations: First, there were only two periods, both without a ranking; second, subjects solved the items only for Domain B. Basing the ranking on pre-sampled performance allows for an online experiment as not all subjects have to participate in the experiment at the same time (for a similar design, see the online experiments in Kirchler et al., 2018). Please note that the ranking was anonymous, private, had no monetary consequences and, thereby, constitutes the mildest form of social comparison, as it potentially addresses self-image concerns but not reputation or status.

As second treatment variable we varied monetary incentives for Domain B. The piece-rate for each correctly solved item in the prosocial task of Domain A was € 10 (donated to UNICEF) across all treatments. The corresponding piece-rate  $y$  in Domain B was € 5 in treatment PR05, € 10 in treatment PR10, and € 15 in treatment PR15. This treatment variation was

implemented between-subjects with random assignment. Thus, combining the between-subject treatments with varying piece-rates, and the within-subject treatments with implementing a ranking or not, we arrive at six treatments outlined in Table 1.

Table 1: Treatment overview: This table outlines treatment details for all treatments. “PR” denotes the individual payment for each solved item, i.e., the piece-rate, in Domain B. The numbers after “PR” indicate the level of the piece-rate in Euro.

	PR10		PR15		PR05	
€ per item in Domain A/B	10/10		10/15		10/5	
Ranking	No	Yes	No	Yes	No	Yes
Periods	2	2	2	2	2	2
Seconds per period	120	120	120	120	120	120
Sessions	within-subjects		within-subjects		within-subjects	
No. Subjects	93		89		104	

After each period subjects received feedback on the number of items solved in Domains A and B. In the corresponding ranking treatments, subjects also received information on their rank in Domain B (see Appendix C for details on the feedback screen).

We included the following self-reported and non-incentivized questionnaire items: First, subjects had to indicate the relevance of the donation to UNICEF (“How important it is for you personally to make an effort to create donations to UNICEF for measles vaccine?”) on a scale from 1 (“not important at all”) to 5 (“very important”) before they started solving APM items. With this question we assessed a subject’s intrinsic motivation for Domain A, i.e., for the prosocial activity.<sup>8</sup> After the experiment, we included the 5-item WOFO questionnaire on competitiveness (Helmreich and Spence, 1978), and the SOEP risk elicitation question on general risk taking (Dohmen et al., 2011).<sup>9</sup> Finally, we also asked for participants’ age, gender and job description.<sup>10</sup>

### 3.2 Implementation of the experiment

The online experiments were run in two waves: In wave 1 we collected data of 12 subjects for the ranking as described above. In the main experiment (wave 2), 286 subjects completed the experiment. Of our participants, 10% were female and the mean age was 37.84 years ( $SD =$

<sup>8</sup>It was posed *before* the experiment to avoid ex post rationalization of behavior and performance in rankings.

<sup>9</sup>The WOFO questions read as follows: “I enjoy working in situations involving competition with others”; “It is important to me to perform better than others on a task”; “I feel that winning is important in both work and games”; “It annoys me when other people perform better than I do”; “I try harder when I am in competition with other people”. Each of these WOFO questions have been answered on a scale from 1 (“strongly disagree”) to 5 (“strongly agree”). The SOEP question read as follows: “How do you see yourself: Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?” and could be answered on a scale from 0 (“risk-averse”) to 10 (“risk-prone”).

<sup>10</sup>For an overview over participants’ self-reported job descriptions, see Table B1 in the Appendix.

8.61). We used proprietary contacts from our BEFORE database (Behavioral Finance Online Research *before.world*) to recruit finance professionals from different EU countries and across a variety of job functions. Since our study is motivated by social comparison and strong status cultures, we chose to go beyond the standard subject pool of students and focus on the finance industry with their prevalent incentive and ranking culture (Kirchler et al., 2018, 2019). Thus, we are confident that running the experiments with finance professionals increases external validity of our results and its interpretation (see also the argument in Kirchler et al. (2018) on the importance of experiments with industry professionals for industry-specific research questions).

We paid out one randomly selected period out of four. Average payment including the participation fee (€10) was €27.26 ( $SD = 20.49$ ;  $Min = 10$ ,  $Max = 130$ ) for a median duration of the experiment of 15 minutes. Hence, with an average hourly wage of more than € 100, we believe that the experiment was well-incentivized.

### 3.3 Hypotheses

Based on the theoretical findings in Section 2, we derive several hypotheses regarding our experimental results. Our analyses is based on two measures: First, TOTAL denotes the total number of correctly solved items in both Domains, A and B, which serves as a proxy for total effort ( $a + b$ ). Second, FRAC(B) denotes the fraction of correct answers in Domain B relative to the total number of correct answers in both Domains, A and B. This is a measure of the fraction of output achieved in Domain B and, thus, serves as a proxy for relative effort put into Domain B.<sup>11</sup>

As indicated by Proposition 1, the impact of an individual’s intrinsic motivation to provide the positive externality (i.e., the relevance of the donation) is straightforward.

**Hypothesis 1** (i) TOTAL is increasing in the self-assessed relevance of the donation; (ii) FRAC(B) is decreasing in the self-assessed relevance of the donation.

Similarly, Proposition 2 directly implies the following hypothesis regarding the impact of piece-rates.

**Hypothesis 2** (i) TOTAL is increasing in the piece-rate; (ii) FRAC(B) is increasing in the piece-rate.

As displayed in Proposition 3, the presence of a ranking may – depending on the shape of the utility from social comparison and ranks achieved in previous periods – introduce an additional incentive or disincentive for effort in Domain B. If we assume that the utility from social comparison  $\Psi$  is monotonically increasing (i.e., a higher rank always induces a positive utility), Corollary 1 implies

---

<sup>11</sup>We chose the number of correct answers instead of the number of tries to solve items as a proxy to measure performance. The reason is that we cannot reasonably distinguish between answers provided with effort and simple tries without any effort provided (“clicking through-behavior”), rendering a variable that focuses on tries noisy.

**Hypothesis 3** (i) TOTAL is increasing in the presence of a ranking; (ii) FRAC(B) is increasing in the presence of a ranking.

If, however, the utility from social comparison is non-monotone, the impact of a ranking may crucially depend on an individual’s rank in previous periods. According to Proposition 3, the higher the effort in Domain B up until period  $t$  (i.e., the better the rank in these periods), the smaller the effort in this domain in period  $t + 1$ . In fact, Proposition 3 demonstrates that a higher effort in Domain B in periods up until  $t$  has the same impact on efforts in period  $t + 1$  as a lower piece-rate  $y$ . Hence, we get the following hypothesis.

**Hypothesis 4** (1) If  $\text{RANK}_1 \dots \text{RANK}_t$  is sufficiently bad: (i) TOTAL is increasing in the presence of a ranking; (ii) FRAC(B) is increasing in the presence of a ranking. (2) If  $\text{RANK}_t$  is sufficiently good: (i) TOTAL is decreasing in the presence of a ranking; (ii) FRAC(B) is decreasing in the presence of a ranking.

As indicated by Propositions 1-3 and Corollary 1, however, the hypothesized impact on TOTAL vanishes as benefits become linear. This means that if the benefit from a ranking tends to be linear (or no ranking is published), the self-assessed relevance of donation has no impact on TOTAL and if the benefit of a donation tends to be linear, piece-rates and the ranking have no impact on TOTAL. In both cases, the linearity of benefits induces a constant rate of substitution between efforts in the two domains. As a result, changing the relevance of donation, the piece-rate, or introducing the ranking only alters the effort division but not total effort spending in this case.

## 4 Results of the experiment

**Descriptive Results** A first overview over the two main variables of interest, TOTAL and FRAC(B) is provided in Figure 2.<sup>12</sup> We do not find major differences in the aggregate numbers for total effort (TOTAL) across treatments. In contrast, the piece-rate seems to have an impact on the fraction of solved items for the selfish activity (Domain B): In Treatment PR15, FRAC(B) is higher than 0.6, exceeding the fractions in Domain B in the other treatments. The ranking provided in treatments of type RANKING shows no overall effect on TOTAL and FRAC(B).

Furthermore, we show the mean of FRAC(B) over all subjects and piece-rates as a function

---

<sup>12</sup>See Figure B1 in the Appendix for mean levels of items solved in Domain A and B, respectively. Note that the variable FRAC(B) is set to 0 in case a subject’s answers are all wrong. Over the two periods this is not the case for any of our subjects, thus the results on Hypotheses 1 and 2 are not affected by this choice. However, in 28 cases subjects failed to answer any of the items correctly in one of the two periods. Reassuringly, our results on Hypotheses 3 and 4 remain robust when dropping these observations (instead of setting FRAC(B) to 0). In addition to analysis below, please also refer to panel regression results reported in Table B3 in the Appendix.

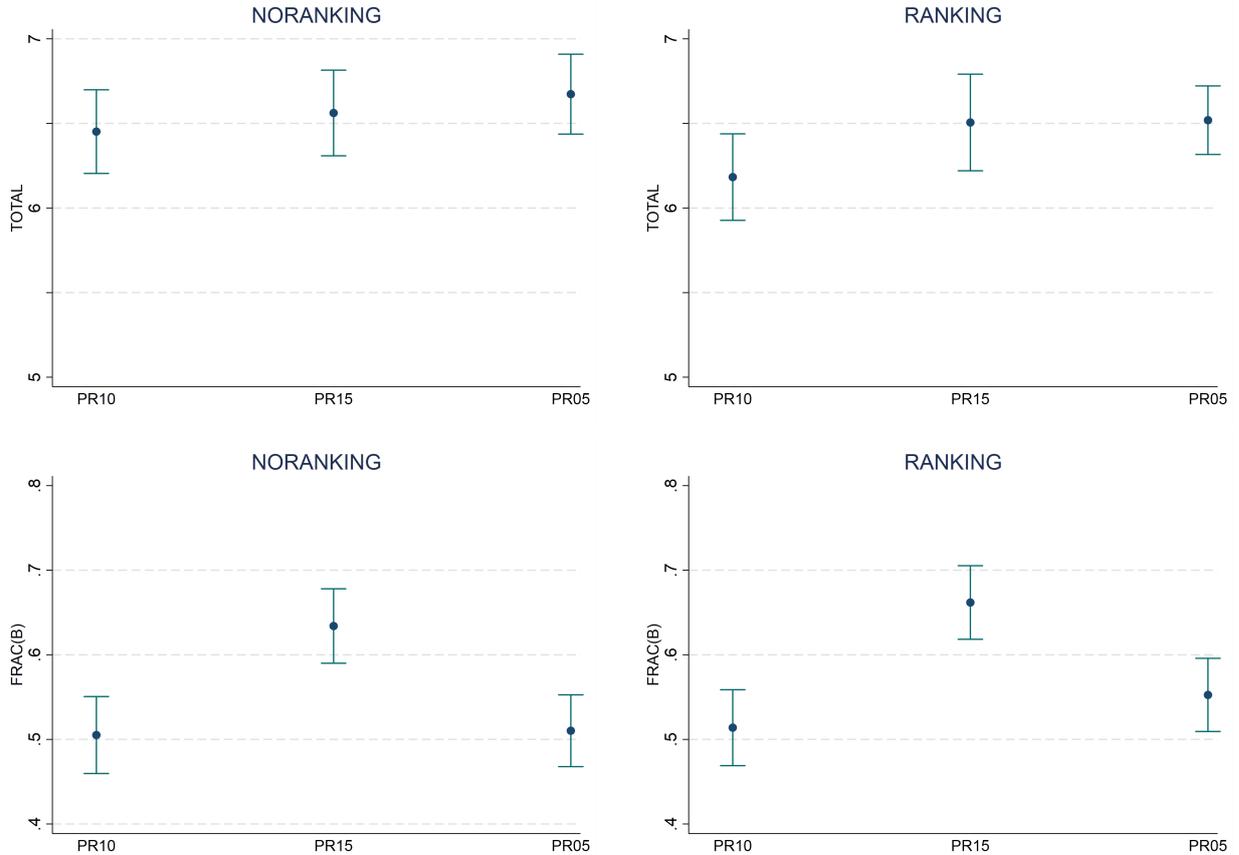


Figure 2: Treatment overview, displaying means (points) and standard errors (bars). Upper panel: TOTAL denotes the total number of solved items in both Domains, A and B, across the two periods. Lower panel: FRAC(B) stands for the fraction of correct items in Domain B (selfish activity) relative to the total number of correct items (right) in both domains. PR10, PR15, and PR05 represent the different treatments with RANKING (right panel) and without (left panel).

of the self-assessed relevance of the donation (DONATION) in Figure 3.<sup>13</sup> We find that the self-assessed relevance of the donation shows high explanatory power in how much effort is provided in each of the two domains. From this figure, one might infer that a substantial fraction of subjects may be refusing to spend effort in the prosocial activity of Domain A or, put differently,

<sup>13</sup>For the sake of comparability across different scales, we standardize the questionnaire variables (with  $ME = 0$  and  $SD = 1$ ). In particular, we subtract the mean from each value and divide it by the standard deviation. For the COMPETITIVENESS questions on competitiveness, we normalize each question separately before computing mean aggregated competitiveness scores which then are also normalized. Please also note that randomization of subjects into treatments resulted in the expected heterogeneity of types in each treatment: Pairwise tests of distributions (Kolmogorov-Smirnov test) of self-reported relevance of donation DONATION do not reveal significant differences between treatments ( $p > 0.05$ ). Distributions of normalized answers are depicted in Figure B2 in the Appendix.

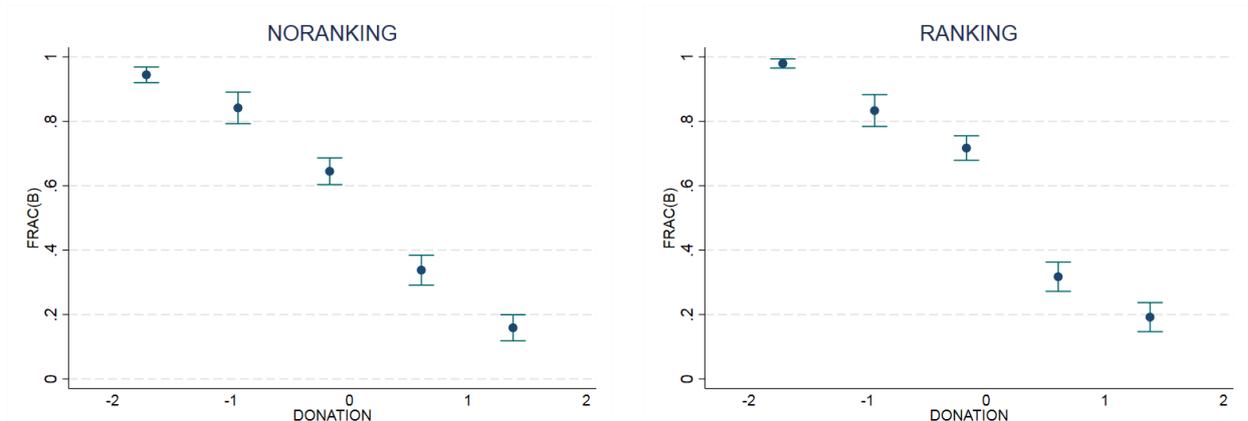


Figure 3: Mean  $\text{FRAC}(B)$ , i.e., the fraction of correct items in Domain B relative to the total number of correct items in both domains, as a function of the self-assessed relevance of the donation ( $\text{DONATION}$ ). The right panel depicts observations from treatments with RANKING and the left panel from treatments without RANKING.

invest all effort into one’s own payment (Domain B). In fact, 31.11% (30.07%) of subjects refuse to spend any effort in Domain B over both periods, i.e.,  $\text{FRAC}(B) = 0$ , in treatments with (without) ranking. In contrast, 39.86% (41.26%) of subjects spend all their effort in Domain B, i.e.,  $\text{FRAC}(B) = 1$ , in treatments without (with) ranking (see also Figure B3 in the Appendix).

**Result Hypothesis 1** The total effort provided ( $\text{TOTAL}$ ) is independent of the self-assessed relevance of the donation, but the relative fraction of effort put in the selfish activity in Domain B ( $\text{FRAC}(B)$ ) decreases with the self-assessed relevance of the donation.

To test Hypothesis 1, we run OLS regressions with  $\text{TOTAL}$  and  $\text{FRAC}(B)$ , respectively, as dependent variables.<sup>14</sup> We add treatment dummies for PR15 and PR05 as explanatory variables (i.e., treatment PR10 serves as baseline), a dummy variable denoting treatments with ranking (RANKING), and variables controlling for self-reported relevance of Domain A ( $\text{DONATION}$ ). We further add CONTROLS in all regressions of the paper except if otherwise noted, including gender, age, self-reported competitiveness, and self-reported risk tolerance. As can be seen from Table 2,  $\text{TOTAL}$  is independent of the self-assessed relevance of donation ( $\text{DONATION}$ ), indicating no support for Hypothesis 1(i). However, in line with Hypothesis 1(ii), we show that the relative effort put in the selfish activity of Domain B ( $\text{FRAC}(B)$ ) decreases with donation relevance ( $\text{DONATION}$ ). The effect size, a 26pp increase for a unit increase in the donation relevance, can be considered substantial. We observe that subjects who considered donations to UNICEF

<sup>14</sup>Please note that the variable  $\text{FRAC}(B)$  is distributed in an interval from 0 and 1. However, for the sake of interpretability, we report linear OLS regression results in the main text and fraction probit regression results in Table B2 in the Appendix. Results are qualitatively robust to the choice of the model.

very important spend on average more than 80 percent of their effort in Domain A, while those considering such donations not important at all spend almost all their effort for Domain B.

**Result Hypothesis 2** TOTAL is independent of the piece-rate, but the effort provided for the selfish activity in Domain B (FRAC(B)) is increasing with a higher piece-rate.

As outlined in Table 2, we find that the piece-rates have no impact on the total effort level (TOTAL). However, a higher piece-rate in treatment PR15 leads to higher relative effort in Domain B (FRAC(B)), indicating a stronger focus on a higher individual payment with increasing monetary incentives. We do not find the reverse effect in treatment PR05. Thus, we can affirmatively answer part (ii) of Hypothesis 2, but not part (i) (for a related finding see Eckartz et al., 2012).

Table 2: OLS regression results with TOTAL and FRAC(B) as dependent variables. TOTAL denotes the total number of correctly solved items in both domains, A and B, across both periods. FRAC(B) stands for the fraction of correctly solved items in Domain B (i.e., the selfish activity) relative to the total number of correctly solved items across both domains and both periods. DONATION indicates the subject’s relevance of the donation (i.e., the prosocial activity of Domain A). PR15 and PR05 are treatment dummies. Treatment PR10 serves as baseline. RANKING indicates a dummy for the ranking treatments. Controls include gender, age, self-reported competitiveness, and self-reported risk tolerance. Robust standard errors are provided in parenthesis. \*  $p < 0.05$ , \*\*  $p < 0.01$ .

	TOTAL	FRAC(B)
PR15	0.145 (0.263)	0.078* (0.035)
PR05	0.324 (0.240)	0.008 (0.031)
RANKING	-0.161 (0.197)	0.027 (0.027)
DONATION	-0.069 (0.089)	-0.264** (0.012)
CONSTANT	7.718** (0.500)	0.715** (0.066)
$N$	572	572
$R^2$	0.048	0.455
$P > F$	0.001	0.000

With regards to Hypotheses 1 and 2, we conclude that the individual preference for the donation and the piece-rate both have an impact on relative, but not on absolute effort. Our findings are particularly noteworthy for conclusions on a substitution effect of monetary incentives: First, with a lower piece-rate we do not observe that subjects decrease their relative effort

in Domain B ( $\text{FRAC}(\text{B})$ ), which might be due to a minimum effort subjects want to provide for their own payment.<sup>15</sup> Overall, we find evidence for the theoretically postulated effects on  $\text{FRAC}(\text{B})$  with an increased piece-rate and conclude that increasing monetary incentives can have a negative impact on prosocial behavior, which is mainly driven by substitution of domains.

**Result Hypothesis 3** The introduction of a ranking does not increase  $\text{TOTAL}$  and  $\text{FRAC}(\text{B})$  on aggregate.

As outlined in Figure 2 and Table 2, the introduction of a ranking does not result in an overall increase in total effort (*abs*) or in a higher relative effort in Domain B ( $\text{FRAC}(\text{B})$ ). In other words, we do not find an aggregate effect of rank incentives, which brings us to the next hypothesis.

**Result Hypothesis 4** Both, the changes in total effort and in the fraction invested in the selfish activity of Domain B from period 1 to period 2 are positively related to the position in the ranking. Those lagging behind in the ranking in period  $t = 1$  increase total effort and put more effort in Domain B in  $t = 2$ . This indicates a “ranking substitution effect” of underperformers, resulting in negative consequences for the prosocial activity. The effect, however, is the opposite for those ahead in the ranking.

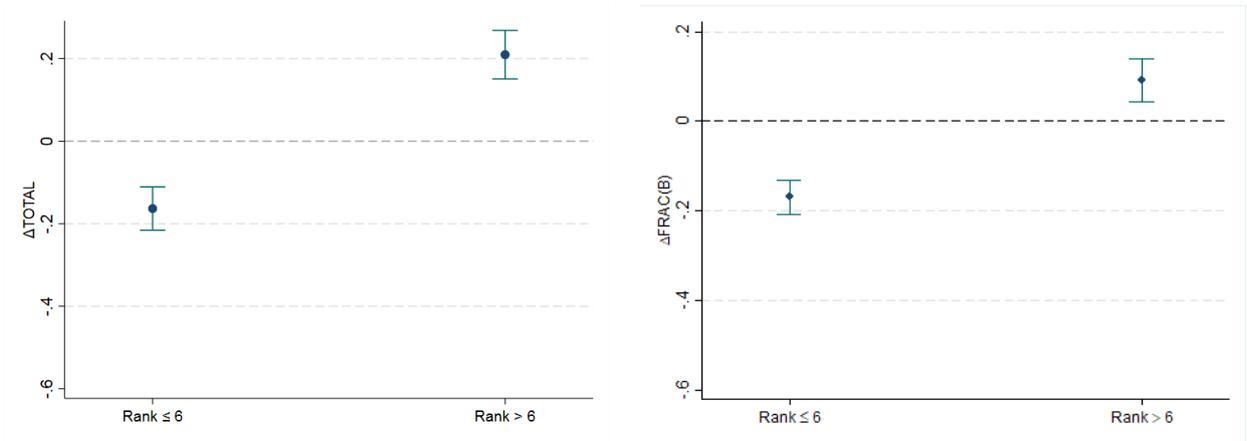


Figure 4: Percentage changes in the number of solved items in total ( $\Delta \text{TOTAL}$ ; left) and in the fraction of solved items in the selfish activity of Domain B ( $\Delta \text{FRAC}(\text{B})$ ; right) from period 1 to period 2 for outperformers (ranks  $\leq 6$ ) and underperformers (ranks  $> 6$ ) in period 1. Points indicate means and bars show standard errors.

To test Hypothesis 4 we examine percentage changes in total effort  $\text{TOTAL}$  and relative effort

<sup>15</sup>See also Figure B3 in the Appendix depicting the fraction of subjects refusing to spend any effort in Domain B, i.e.,  $\text{FRAC}(\text{B}) = 0$ , which is higher in treatment PR15, but unchanged in treatment PR05 compared to treatment PR10.

in Domain B  $\text{FRAC}(\text{B})$  from period 1 to period 2 given the observed rank in the first period:  $\Delta\text{TOTAL} = \frac{\text{TOTAL}_{t=2} - \text{TOTAL}_{t=1}}{\text{TOTAL}_{t=1}}$  and  $\Delta\text{FRAC}(\text{B}) = \frac{\text{FRAC}(\text{B})_{t=2} - \text{FRAC}(\text{B})_{t=1}}{\text{FRAC}(\text{B})_{t=1}}$ .<sup>16</sup> Importantly, since in our experiment, subjects faced a random sequence of APM items, it may be that periods 1 and 2 vary in the difficulty of the items. Hence, for calculating  $\Delta\text{TOTAL}$ , we adjust variable  $\text{TOTAL}_t$  by normalizing it with the average number of correct items over all participants in the same period in the same treatment for the same random sequence of APM items.  $\Delta\text{TOTAL}$  measures the change in total effort from period 1 to period 2 and can be interpreted as above- and below-average changes in total effort. Since the variable  $\text{FRAC}(\text{B})$  captures the fraction of effort spent in Domain B for each subject, no such adjustment for the difficulty of items is necessary. Thus, it is important to keep in mind that the variables  $\text{TOTAL}$  and, in turn,  $\Delta\text{TOTAL}$  are calculated differently when testing Hypothesis 4 compared to the rest of the paper.

Figure 4 provides first evidence on Hypothesis 4. We observe that those at the top of the ranking reduce total effort and decrease the fraction of solved items in Domain B. Instead, those lagging behind do exactly the opposite. They increase total effort and switch more frequently to the selfish activity of Domain B, indicating a ranking substitution effect.

To test Hypothesis 4, we run OLS regressions with  $\Delta\text{TOTAL}$  and  $\Delta\text{FRAC}(\text{B})$ , respectively, serving as dependent variables (see Table 3). The main explanatory variable of interest is the binary variable  $\text{UNDERPERFORM}$ , which takes the value 1 in case a subject’s rank is below average ( $> 6$ ) and 0 otherwise.<sup>17</sup> With this variable we can examine whether “out- and underperformers” react differently to the observed performance. To control for period effects that cannot be attributed to the ranking, we also include variable  $\Delta\text{NORANKING}$ , which denotes values of  $\text{TOTAL}$  and  $\text{FRAC}(\text{B})$ , respectively, of the corresponding treatments without ranking. This variable captures non-ranking related effects of  $\Delta\text{TOTAL}$  and  $\Delta\text{FRAC}(\text{B})$  from one period to the next – for instance due to moral licensing, regret, or due to feedback effects.

We find statistical support for Hypothesis 4, as subjects strongly react to the ranking relative to their peers presented in period 1. Those with a bad relative performance in period 1 increase total effort and substitute between both domains by increasing the fraction solved in the selfish activity of Domain B at the expense of the prosocial activity – and *vice versa* for the outperforming peers.<sup>18</sup>

---

<sup>16</sup>Please note, if  $\text{TOTAL}_{t=1} = 0$  then  $\Delta\text{TOTAL} = \text{TOTAL}_{t=2}$  and if  $\text{FRAC}(\text{B})_{t=1} = 0$ , then  $\Delta\text{FRAC}(\text{B}) = \text{FRAC}(\text{B})_{t=2}$ .

<sup>17</sup>Please note that we chose the split in above- and below average performance, because subjects compete against a pre-sampled selection of peers. Therefore, for an individual subject the expected value that splits above- and below-average performance are ranks 6 and 7. Results on an specification of OLS regressions with the individual’s rank ( $\text{RANK}_{t=1}$ ) as explanatory variable instead of  $\text{UNDERPERFORM}$  are reported in Table B4 in the Appendix. Results remain qualitatively similar (see also Figure B4 in the Appendix).

<sup>18</sup>See Figure B5 in the Appendix for further evidence that the observed effect can be attributed to ranking. Figure B5 is similar to Figure 4 above, but for observations in treatments  $\text{NORANKING}$ . One can see that there is no difference in change in behavior, neither with regards to  $\text{TOTAL}$  nor  $\text{FRAC}(\text{B})$ , between out- and underperformer. Note that separation in out- and underperformer is made according to ranks from treatments  $\text{RANKING}$  in order to compare the same subjects across treatments.

Table 3: OLS regression results with the following dependent variables: percentage changes in the number of solved items ( $\Delta_{\text{TOTAL}}$ ) and changes in the fraction of solved items in the selfish activity of Domain B ( $\Delta_{\text{FRAC(B)}}$ ) from period 1 to period 2. Variable UNDERPERFORM indicates a binary dummy variable for underperforming subjects in period 1 ( $\text{Rank} > 6$ ).  $\Delta_{\text{NORANKING}}$  indicates the changes as described in the respective treatments without ranking. DONATION stands for the self-reported relevance of the donation. PR15 and PR05 indicate treatment dummies. Treatment PR10 serves as baseline. Controls include gender, age, self-reported competitiveness, and self-reported risk tolerance. Robust standard errors are provided in parenthesis. \*  $p < 0.05$ , \*\*  $p < 0.01$ .

	$\Delta_{\text{TOTAL}}$	$\Delta_{\text{FRAC(B)}}$
UNDERPERFORM	0.410** (0.093)	0.319** (0.074)
$\Delta_{\text{NORANKING}}$	-0.131 (0.074)	0.028 (0.117)
PR15	0.087 (0.125)	0.113 (0.101)
PR05	0.128 (0.101)	0.116 (0.080)
DONATION	-0.021 (0.044)	-0.050 (0.031)
CONSTANT	-0.049 (0.221)	-0.296 (0.182)
$N$	286	286
$R^2$	0.077	0.054
$P > F$	< 0.000	0.016

To sum up, we conclude that utility gained from social comparison is not monotone and that the reaction to observed ranks can mainly be explained by a ranking substitution effect between Domain A and Domain B. Those achieving a good rank decrease their total effort as well as their *relative* effort for the selfish domain. In contrast, those lagging behind increase total effort and, too, substitute between both domains by decreasing the number of solved items for the prosocial activity in Domain A, and increasing their relative effort in the selfish activity of Domain B.

Additionally and in more exploratory terms, we examine which characteristics explain a subject's rank. We run ordered logistic regressions using a subject's rank in a period as dependent variable ( $\text{RANK}_{t=n} \in \{1, 2, 7, 9, 12\}$  with  $n$  indicating period 1 or 2). As explanatory variables we include a subjects' self-assessed answers regarding donation relevance, competitiveness, and risk tolerance (DONATION, COMPETITIVENESS, and RISK), as well as a gender dummy (FEMALE) and age (AGE). For the regression on the individuals' rank in period 2 (model 2) we also include  $\text{RANK}_{t=1}$ , i.e., the observed rank in period 1.

Table 4: Ordered logistic regression results with a subject’s rank in period  $n$  ( $\text{RANK}_{t=n}$ , ranging from 1 to 12 with  $n$  indicating period 1 or 2) as dependent variable.  $\text{RANK}_{t=1}$  is included as explanatory variable in the specification for period 2 indicating a subjects’ rank in period 1.  $\text{DONATION}$  stands for a subject’s relevance of the donation,  $\text{COMPETITIVENESS}$  measures an individual’s competitiveness, and  $\text{RISK}$  indicates self-reported risk tolerance. Gender ( $\text{FEMALE}$ ) and age ( $\text{AGE}$ ) measure additional explanatory variables. Coefficients are reported as odds ratios. Robust standard errors are provided in parenthesis. \*  $p < 0.05$ , \*\*  $p < 0.01$ .

	<b>RANK<sub>t=1</sub></b>	<b>RANK<sub>t=2</sub></b>
RANK <sub>t=1</sub>		1.161** (0.041)
COMPETITIVENESS	0.777* (0.092)	1.043 (0.149)
DONATION	3.059** (0.411)	1.788** (0.294)
RISK	0.909 (0.113)	0.964 (0.133)
FEMALE	0.879 (0.363)	0.667 (0.342)
AGE	1.026 (0.015)	1.021 (0.013)
$N$	286	286
$PseudoR^2$	0.121	0.117
$P > F$	0.000	0.000

Results of the regression analyses are reported in Table 4. Please note that coefficients are reported as odds ratios. We find that the observed rank in period 1 explains a subject’s rank in period 2. Together with the results described above, this indicates that, on average, subjects ahead in the ranking seem to manage to keep their good relative position despite the reduced effort in Domain B. Moreover, we report that in the first period subjects’ self-assessed competitiveness,  $\text{COMPETITIVENESS}$ , explains their rank, with more competitive subjects achieving a better rank. This relation becomes insignificant once the ranking is announced in period 2. Potentially, this is an indication of “crowding out” of intrinsic levels of competitiveness by the externally provided ranking. Finally, we find that higher self-reported donation relevance ( $\text{DONATION}$ ) leads to a worse rank, as more effort is invested in the prosocial activity of Domain A. The effect size is considerable, even though it decreases in the second period (while remaining significant).

## 5 Discussion

Social comparison based on rankings features most prominently in labor markets with strong competition for talent (e.g., financial professionals, science, top managers etc.). One objective of rankings is information provision as it allows (potential) employers and/or peers to identify top-performers. Moreover, rankings are also often used as an incentive device: Payments are explicitly or implicitly contingent on the rank and the goal to achieve a high rank may also be a motive for better performance *per se*. Our results, however, indicate some hidden costs of the introduction of rankings and potential lessons to be learned for the design of incentive schemes.

### 5.1 Hidden Costs of Rankings

In contrast to previous studies on the impact of piece-rates and rankings in single-task environments, our investigation of a multi-tasking environment identifies two yet unrecognized effects:

*Ranking Substitution Effect:* Neither an exogenous variation of piece-rates nor the considerable heterogeneity of individual’s self-reported assessment of the relevance of the donation have a significant impact on total effort in our experiment. However, participants *substitute* efforts spent in the two domains. The higher the piece-rate or the lower the relevance of the donation, the higher is the fraction of effort spent in the domain that determines the agent’s earnings (Domain B). While this does not imply that more substantial piece-rate variations may not also have an impact on total effort, it clearly indicates that any attempt to increase total effort with higher piece-rates in Domain B would at least come at the cost of significant substitution of efforts across domains.

In contrast to piece-rates, a ranking neither influences total nor relative effort on the aggregate level in our experiment. This is because the impact of a ranking on an individual’s decision crucially depends on the individual’s rank. This is captured by the second effect.

*Ranking Attenuation Effect:* The impact of a ranking on an individual with a bad rank is similar to an enhanced piece-rate: The individual spends more effort in Domain B (and also increases total effort). In contrast, the impact of a ranking on an individual with a good rank is similar to a reduced piece-rate: The individual spends less effort in Domain B (and also reduces total effort). As a potential result, the difference of total efforts and relative efforts might diminish in response to a ranking. This is in stark contrast to the impact of rankings in single-task environments (without a prosocial dimension) where a ranking is often regarded as an additional incentive to spend effort and, accordingly, has the same impact as an enhanced piece-rate (see, however, the inconclusive result in the literature reported in Section 1). In this case, more productive agents (i.e., agents with lower effort costs or higher returns to a given piece-rate) spend *more* effort compared to less productive agents in the presence of a ranking which yields a larger difference in total and relative effort in response to a ranking.

## 5.2 Implications for Contract Design

The substitution and attenuation effects have several implications for the optimal design of incentive schemes.

1) *Agency Costs*: For single-task environments it is often assumed that a given effort by agents can be achieved for a lower piece-rate (i.e., at lower costs) in the presence of a ranking. Our findings for a multi-tasking environment with a prosocial dimension indicate that this is only true for agents lagging behind in the ranking. For outperforming agents, the ranking rather operates similar to a piece-rate reduction. Hence, lowering the piece-rate and introducing a ranking may lead to the same total and relative effort by underperforming agents, but introduces a disincentive for total and relative effort for agents leading in the ranking. As a consequence, the introduction of a ranking does not unambiguously reduce agency costs – it might work as a rather imperfect substitute for piece-rates.

2) *Bonus Caps*: Receiving similar efforts for reduced piece-rates in the presence of a ranking is not only reducing agency costs, it may also be regarded as a valuable tool for maintaining incentives in the presence of payment-regulations. Consider, for example, the frequently discussed cap of bonus payments (see, e.g., Bénabou and Tirole, 2016): If piece-rates are regulated to be below a certain threshold value, a ranking would be a tool to further increase total and relative effort into the task that is subject to the ranking. Our findings suggest that offsetting (regulated) piece-rates by a ranking may induce significant substitution across tasks and disincentives for outperforming agents. If these substitution and attenuation effects are sufficiently pronounced, introducing the ranking may be inferior not only from a welfare perspective, but also from the firm’s point of view.

3) *Selecting productive agents*: In a single-task environment, it is often assumed that introducing a ranking enhances total efforts and increases the difference between efforts spent by more or less productive and/or intrinsically motivated agents for a given piece-rate. As a consequence, offering contract menus tailored to more or less productive agents becomes more attractive in the presence of a ranking. In contrast, substitution and attenuation in a multi-tasking environment reduces the effect of a ranking on total effort and potentially on the difference between total and relative efforts by agents who lead or lag behind in the ranking. This makes the selection of agents with different productivity and/or intrinsic motivation not only less attractive but also less feasible. Ulrichshofer and Walzl (2020) add utility from social comparison as in Section 2 to the set-up by Bénabou and Tirole (2016) and analyze how substitution and attenuation alters optimal screening contracts depending on the intensity of labor market competition. If employers have a lot of market power, a psychological cost from lagging behind in a ranking reduces efficiency distortions of optimal screening contracts, as contracts with a low piece-rate become less attractive for high productivity agents. If, however, the competition intensity is high, psychological costs of lagging behind only amplify efficiency distortions, as contracts with

a high piece-rate become more attractive for low productivity agents. In contrast – and for the same reason – psychological costs of leading in the ranking enhance efficiency in this case.

## 6 Conclusion

In this paper, we developed a theoretical multi-tasking model of an agent that derives utility from monetary payments and social comparison, but also from a prosocial activity. In particular, we examined how monetary and rank incentives influence effort in this theoretical environment. Depending on how the agent interprets the ranking, it can serve as an additional incentive or disincentive for the monetarily incentivized task. If a good rank is desirable, it increases total effort and the fraction of effort spent for the ranked task. If a good rank is costly because it is regarded as a signal (to the self or others) of low prosocial activity, it lowers total effort and the fraction of effort spent into the ranked domain. If the agent balances the conflicting motives over time, we expect a ranking to be a disincentive for the ranked task if the agent’s performance was ranked highly in previous periods and an incentive for the ranked task in case of a poor ranking in previous periods.

We tested the hypotheses in a controlled online experiment with 286 internationally operating finance professionals. We found that the introduction of a ranking on the monetarily incentivized activity leads to a “ranking substitution effect”: Outperforming professionals substantially reduced total effort and substituted relative effort spent for their own payment by putting more effort in the prosocial activity. In contrast, underperforming professionals increased total effort and substituted by spending more effort for their own payment and less for the prosocial activity. Moreover, we found evidence for another substitution effect, as higher monetary incentives lead to less effort provided for the prosocial activity and more for the own payment.

We expect our results to be relevant outside the lab for the following reasons: First, since we found ranking effects in anonymous online experiments, we can think of even more pronounced effects in some real life settings where (public) status plays an important role and can effect monetary incentives, too. In addition, in our experiment it is public knowledge that every participant is a finance professional. Given that social comparison is stronger among peers, we expect our experimental design to allow for observing results with rather high external validity. Second, in the business world, there are plenty of situations in which individuals are already at the maximum of expanding effort. A working day is limited with individuals deciding on which activities to focus on the most. In our experiment, subjects also have a limited time budget of two minutes to solve as many items as possible and thereby we believe to mimic a crucial real-world feature with our model.

Finally, given that our findings indicate hidden costs of the introduction of rankings and potential lessons to be learned for the design of incentive schemes, further research on the interplay of monetary incentives, social comparison, and prosocial behavior in multi-tasking problems is

needed. Since our findings are relevant in theory and in the lab, testing our hypotheses in the field would be a fruitful avenue of future research.

## References

- Al-Ubaydli, Omar, John A. List. 2015. On the generalizability of experimental results in economics. G. Frechette, A. Schotter, eds., *Handbook of Experimental Economic Methodology*. Oxford University Press, 420–463.
- Anderson, Melissa, Emily Ronning, Raymond De Vries, Brian Martinson. 2007. The perverse effects of competition on scientists' work and relationships. *Science and Engineering Ethics* **13**(4) 437–461.
- Ariely, Dan, Anat Bracha, Stephan Meier. 2009. Doing good or doing well? image motivation and monetary incentives in behaving prosocially. *American Economic Review* **99** 544–555.
- Azmat, Ghazala, Nagore Iriberry. 2010. The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics* **94**(7-8) 435–452.
- Baker, George. 1992. Incentive contracts and performance measurement. *Journal of Political Economy* **100**(3) 598–614.
- Ball, Sheryl, Catherine Eckel, Philip Grossman, William Zame. 2001. Status in markets. *Quarterly Journal of Economics* **116**(1) 161–188.
- Bandiera, Oriana, Iwan Barankay, Imran Rasul. 2010. Social incentives in the workplace. *The Review of Economic Studies* **77** 417–458.
- Barankay, Iwan. 2015. Rank incentives: Evidence from a randomized workplace experiment. *Working Paper* .
- Bénabou, Roland, Jean Tirole. 2006. Incentives and prosocial behavior. *American Economic Review* **96**(5) 1652–1678.
- Bénabou, Roland, Jean Tirole. 2016. Bonus culture: competitive pay, screening, and multitasking. *Journal of Political Economy* **124**(2) 305–370.
- Blanes-i-Vidal, Jordi, Mareike Nossol. 2011. Tournaments without prizes: Evidence from personnel records. *Management Science* **57**(10) 1721–1736.
- Bolton, Gary, Axel Ockenfels. 2000. Erc: A theory of equity, reciprocity, and competition. *American Economic Review* **90** 166–193.
- Borghans, Lex, Huub Meijers, Bas ter Weel. 2013. The importance of intrinsic and extrinsic motivation for measuring iq. *Economics of Education Review* **34** 17–28.

- Bowles, Samuel, Sandra Polania-Reyes. 2012. Economic incentives and social preferences: Substitutes or complements? *Journal of Economic Literature* **50**(2) 368–425.
- Brown, Keith C., W.V. Harlow, Laura T. Starks. 1996. Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry. *Journal of Finance* **51**(1) 85–110.
- Charness, Gary, David Masclet, Marie Claire Villeval. 2014. The dark side of competition for status. *Management Science* **60**(1) 38–55.
- Cohn, Alain, Ernst Fehr, Lorenz Götte. 2015. Fair wages and effort provision: combining evidence from a choice experiment and a field experiment. *Management Science* 1777–1794.
- Coles, Jeffrey L, Zhichuan (Frank) Li, Albert Y Wang. 2018. Industry tournament incentives. *The Review of Financial Studies* **31**(4) 1418–1459.
- Diamond, Douglas, Raghuram Rajan. 2009. The credit crisis: Conjectures about causes and remedies. *American Economic Review* **99**(2) 606–610.
- Dijk, Oege, Martin Holmen, Michael Kirchler. 2014. Rank matters – the impact of social competition on portfolio choice. *European Economic Review* **66** 97–110.
- Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, Gert Wagner. 2011. Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association* **9**(3) 522–550.
- Eckartz, Katharina, Oliver Kirchkamp, Daniel Schunk. 2012. How do incentives affect creativity? *Working Paper* .
- Falk, Armin, Nora Szech. 2019. Competing image concerns: Pleasures of skill and moral values. *Working Paper* .
- Fanelli, Daniele. 2010. Do pressures to publish increase scientists’ bias? an empirical support from us states data. *PLOS ONE* **5** 1–7.
- Fehr, Ernst, Klaus Schmidt. 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, **115** 817–868.
- Festinger, Leon. 1954. A theory of social comparison processes. *Human Relations* **7** 117–140.
- Fishbach, Ayelet, Ying Zhang, Mingjung Koo. 2009. The dynamics of self-regulation. *European Review of Social Psychology* **20** 315–344.
- Gill, David, Zdenka Kissova, Jaesun Lee, Victoria Prowse. 2019. First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision. *Management Science* **65**(2) 494–507.

- Gneezy, Uri, John List, Jeffrey Livingston, Xiangdong Qin, Sally Sadoff, Yang Xu. 2019. Measuring success in education: The role of effort on the test itself. *American Economic Review: Insights* **1**(3) 291–308.
- Gneezy, Uri, Aldo Rustichini. 2000a. A fine is a price. *The Journal of Legal Studies* **29**(1) 1–17.
- Gneezy, Uri, Aldo Rustichini. 2000b. Pay enough or don't pay at all. *The Quarterly Journal of Economics* **115**(3) 791–810.
- Harrison, Glenn, John List. 2004. Field experiments. *Journal of Economic Literature* **42** 1009–1055.
- Helmreich, Robert L., Janet T. Spence. 1978. The work and family orientation questionnaire: An objective instrument to assess components of achievement motivation and attitudes toward family and career. *Catalog of Selected Documents in Psychology* **8**(2) Document 1677.
- Kaniel, Ron, Robert Parham. 2017. WSJ category kings – The impact of media attention on consumer and mutual fund investment decisions. *Journal of Financial Economics* **123**(2) 337–356.
- Kirchler, M., F. Lindner, U. Weitzel. 2019. Delegated decision making and social competition in the finance industry. *Working Papers in Economics and Statistics, University of Innsbruck* .
- Kirchler, Michael, Jürgen Huber, Matthias Stefan, Matthias Sutter. 2016. Market design and moral behavior. *Management Science* **62**(9) 2615–2625.
- Kirchler, Michael, Florian Lindner, Utz Weitzel. 2018. Rankings and risk-taking in the finance industry. *The Journal of Finance* **73**(5) 2271–2302.
- Kuhnen, Camelia, Agnieszka Tymula. 2012. Feedback, self-esteem, and performance in organizations. *Management Science* **58**(1) 94–113.
- Kuziemko, Ilyana, Ryan Buell, Taly Reich, Michael I. Norton. 2014. “last-place aversion”: Evidence and redistributive implications. *Quarterly Journal of Economics* **129**(1) 105–149.
- Mas, Alexandre, Enrico Moretti. 2009. Peers at work. *American Economic Review* **99**(1) 112–145.
- Mellström, Carl, Magnus Johannesson. 2008. Crowding out in blood donation: Was titmuss right? *Journal of the European Economic Association* **6**(4) 845–863.
- Moldovanu, Benny, Aner Sela, Xianwen Shi. 2007. Contests for status. *Journal of Political Economy* **115**(2) 338–363.
- Osterloh, Margit, Bruno Frey. 2015. Ranking games. *Evaluation Review* **39**(1) 102–129.

- Rajan, Raghuram G. 2006. Has finance made the world riskier? *European Financial Management* **12**(4) 499–533.
- Raven, John. 2000. The Raven’s progressive matrices: Change and stability over culture and time. *Cognitive Psychology* **41**(1) 1–48.
- Shleifer, Andrei. 2004. Does competition destroy ethical behavior? *The American Economic Review* **94** 414–418.
- Sirri, Erik R., Peter Tufano. 1998. Costly search and mutual fund flows. *Journal of Finance* **53**(5) 1589–1622.
- Tran, Anh, Richard Zeckhauser. 2012. Rank as an inherent incentive: Evidence from a field experiment. *Journal of Public Economics* **96** 645–650.
- Ulrichshofer, Anna, Markus Walzl. 2020. Social comparison and optimal contracts in the competition for managerial talent. *mimeo* .

## Appendix

### A Proofs

In period  $t$ , agent  $i$  chooses  $a_t^*$  and  $b_t^*$  to maximize

$$U_i(a(t), b(t); y, z) = \alpha\Phi\left(\sum_{\tau=1 \dots t} a_\tau\right) + \beta\Psi\left(\sum_{\tau=1 \dots t} R(b_\tau)\right) + b_t y + z - \sum_{\tau=1 \dots t} \frac{(a_\tau + b_\tau)^2}{2}.$$

For further reference, we introduce the following notation for the agent's first order condition:  $F_{a_t} = \frac{\partial U_i}{\partial a_t}$  and  $F_{b_t} = \frac{\partial U_i}{\partial b_t}$ . To analyze comparative statics,  $F_{a_t}$  and  $F_{b_t}$  are used as implicit functions, i.e.,  $F_{a_t}(a_t^*, b_t^*, x) = 0$  and  $F_{b_t}(a_t^*, b_t^*, x) = 0$  for choice variables  $a_t^*$  and  $b_t^*$  and parameters  $x \in \{\alpha, \beta, y\}$ . We will omit the subscript  $t$  and superscript  $*$  in  $a_t^*$  and  $b_t^*$  whenever no misunderstanding can arise, i.e., if we consider optimal efforts in a fixed period. Applying the implicit function theorem, fixes the dependence of efforts  $a$  and  $b$  on parameter  $x$  to be

$$\frac{da}{dx} = -\frac{\frac{\partial F_a}{\partial x} \frac{\partial F_b}{\partial b} - \frac{\partial F_b}{\partial x} \frac{\partial F_a}{\partial b}}{\frac{\partial F_a}{\partial a} \frac{\partial F_b}{\partial b} - \frac{\partial F_b}{\partial a} \frac{\partial F_a}{\partial b}}.$$

(we get  $\frac{db}{dx}$  permuting  $a$  and  $b$ ) By the convexity of costs, the weak concavity of  $\Phi$  and  $\Psi$  (and the monotonicity of  $R$ ), we get  $\frac{\partial F_a}{\partial a} < 0$  and  $\frac{\partial F_b}{\partial b} < 0$ . As effort costs are  $(a+b)^2/2$ , we get  $\frac{\partial F_a}{\partial b} = -1$  and  $\frac{\partial F_b}{\partial a} = -1$ . This guarantees that the denominator in  $\frac{da}{dx}$  is positive. In the remainder of this appendix we will refer to the denominator as  $(\cdot)$ .

#### Proof of Proposition 1:

- (i) As  $\frac{\partial F_b}{\partial \alpha} = 0$  and  $\frac{\partial F_a}{\partial \alpha} = \frac{\partial \Phi}{\partial \alpha} > 0$ ,  $\frac{da}{d\alpha} = -\frac{\frac{\partial F_a}{\partial \alpha} \frac{\partial F_b}{\partial b}}{(\cdot)} > 0$ .
- (ii) As  $\frac{\partial F_b}{\partial \alpha} = 0$  and  $\frac{\partial F_a}{\partial \alpha} = \frac{\partial \Phi}{\partial \alpha} > 0$ ,  $\frac{db}{d\alpha} = -\frac{\frac{\partial F_a}{\partial \alpha}}{(\cdot)} < 0$ .
- (iii) With (i) and (ii),  $\frac{d(a+b)}{d\alpha} = -\frac{\frac{\partial \Phi}{\partial \alpha} \beta \frac{\partial^2 \Psi}{\partial b^2}}{(\cdot)} > 0$  if  $\Psi$  is strictly concave (and  $= 0$  if  $\Psi$  is linear).
- (iv) With (i) - (iii),

$$\frac{db/(a+b)}{d\alpha} = \frac{\frac{db}{d\alpha}(a+b) - b \frac{d(a+b)}{d\alpha}}{(a+b)^2} = \frac{\frac{\partial \Phi}{\partial \alpha}}{(a+b)^2 \cdot (\cdot)} (b\beta \frac{\partial^2 \Psi}{\partial b^2} - (a+b)) < 0.$$

#### Proof of Proposition 2:

- (i) As  $\frac{\partial F_a}{\partial y} = 0$  and  $\frac{\partial F_b}{\partial y} = 1$ ,  $\frac{da}{dy} = -\frac{\frac{\partial F_b}{\partial y}}{(\cdot)} < 0$ .
- (ii) As  $\frac{\partial F_a}{\partial y} = 0$  and  $\frac{\partial F_b}{\partial y} = 1$ ,  $\frac{db}{dy} = -\frac{\frac{\partial F_b}{\partial y} \frac{\partial F_a}{\partial a}}{(\cdot)} > 0$ .

(iii) With (i) and (ii),  $\frac{d(a+b)}{dy} = -\frac{\alpha \frac{\partial^2 \Phi}{\partial a^2}}{()} > 0$  if  $\Phi$  is strictly concave (and  $= 0$  if  $\Phi$  is linear).

(iv) With (i) - (iii),

$$\frac{db/(a+b)}{dy} = \frac{\frac{db}{dy}(a+b) - b \frac{d(a+b)}{dy}}{(a+b)^2} = -\frac{1}{(a+b)^2 \cdot ()} (a(\alpha \frac{\partial^2 \Phi}{\partial a^2} - 1) - b) > 0.$$

**Proof of Corollary 1 and Proposition 3** By the definition of  $R^*$ ,  $\frac{\partial \Psi}{\partial b} > 0$  iff  $\sum_{\tau=1\dots t} R(b(\tau)) < R^*$  and  $\frac{\partial \Psi}{\partial b} < 0$  iff  $\sum_{\tau=1\dots t} R(b(\tau)) > R^*$ . Then, the Corollary and the Proposition follow from the following observations:

(i) As  $\frac{\partial F_a}{\partial \beta} = 0$  and  $\frac{\partial F_b}{\partial \beta} = \frac{\partial \Psi}{\partial b}$ ,  $\frac{da}{d\beta} = -\frac{\frac{\partial F_b}{\partial \beta}}{()}$ .

(ii) As  $\frac{\partial F_a}{\partial \beta} = 0$  and  $\frac{\partial F_b}{\partial \beta} = \frac{\partial \Psi}{\partial b}$ ,  $\frac{db}{d\beta} = -\frac{\frac{\partial F_a}{\partial a} \frac{\partial F_b}{\partial \beta}}{()} = -\frac{\frac{\partial F_a}{\partial a} \frac{\partial \Psi}{\partial b}}{()}.$

(iii) With (i) and (ii),  $\frac{d(a+b)}{d\beta} = -\frac{\frac{\partial \Psi}{\partial b} \alpha \frac{\partial^2 \Phi}{\partial a^2}}{()}.$

(iv) With (i) - (iii),

$$\frac{db/(a+b)}{d\beta} = \frac{\frac{db}{d\beta}(a+b) - b \frac{d(a+b)}{d\beta}}{(a+b)^2} = -\frac{\frac{\partial \Psi}{\partial b}}{(a+b)^2 \cdot ()} (a(\alpha \frac{\partial^2 \Phi}{\partial a^2} - 1) - b).$$

## B Additional figures and tables

Table B1: Overview over Participants – self-reported variables

	Mean/Percentage
Age	37.84
Female	10.14%
Portfolio-, Fund-, or Asset-Management	40.91%
Trading/Brokerage	6.64%
Financial Advise	7.34%
Investment Banking	7.34%
Analysis/Research	13.64%
Sales	4.20%
Private Banking	3.85%
Risk Management	7.34%
Other	8.74%

Table B2: Fraction probit regression results with  $\text{FRAC}(\mathbf{B})$ , i.e., the fraction of correctly solved items in Domain B relative to the total number of correctly solved items across both domains and both periods, as dependent variables.  $\text{DONATION}$  indicates the subject's self-reported relevance of the donation.  $\text{PR15}$  and  $\text{PR05}$  are treatment dummies.  $\text{RANKING}$  indicates a dummy for the ranking treatments. Treatment  $\text{PR10}$  serves as baseline. Controls include gender, age, self-reported competitiveness, and self-reported risk tolerance. Robust standard errors are provided in parenthesis. \*  $p < 0.05$ , \*\*  $p < 0.01$ .

	$\text{FRAC}(\mathbf{B})$
$\text{PR15}$	0.256* (0.129)
$\text{PR05}$	-0.007 (0.108)
$\text{RANKING}$	0.101 (0.096)
$\text{DONATION}$	-0.883** (0.057)
$\text{CONSTANT}$	0.738** (0.245)
$N$	572
$PseudoR^2$	0.293
$P > F$	0.000

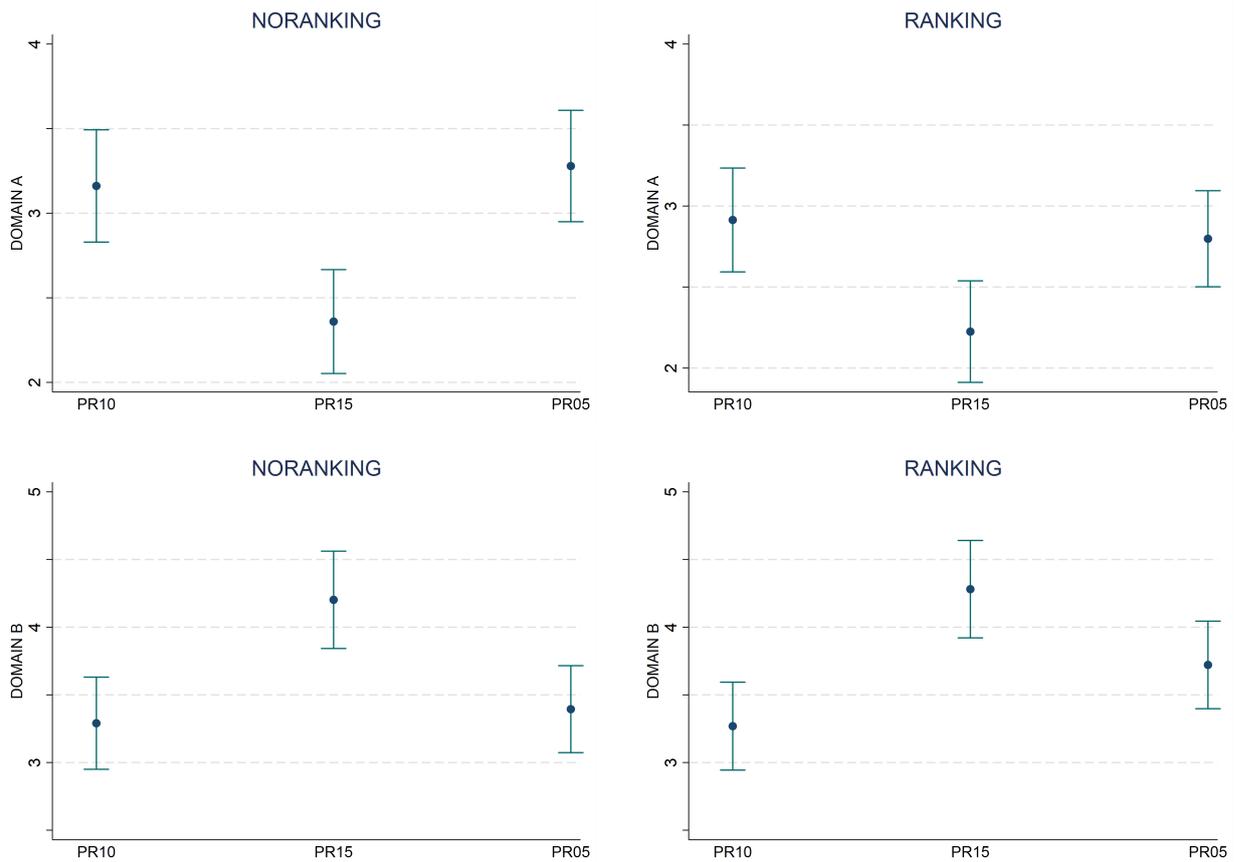


Figure B1: Treatment overview of items solved in Domain A and B, respectively. Means are indicated by points, standard errors by bars. PR10, PR15, and PR05 represent the different treatments with RANKING (right panel) and without (left panel).

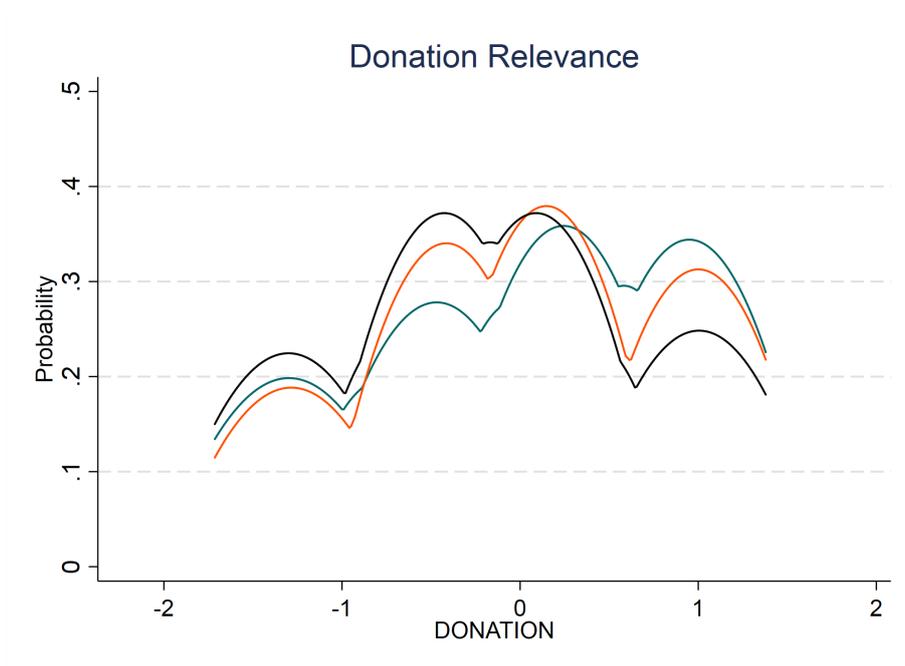


Figure B2: Distribution of the normalized variable for self-reported relevance of donation to UNICEF (DONATION) for each treatment. Treatment PR10 is depicted by blue lines, PR15 by orange lines, and PR05 by black lines.

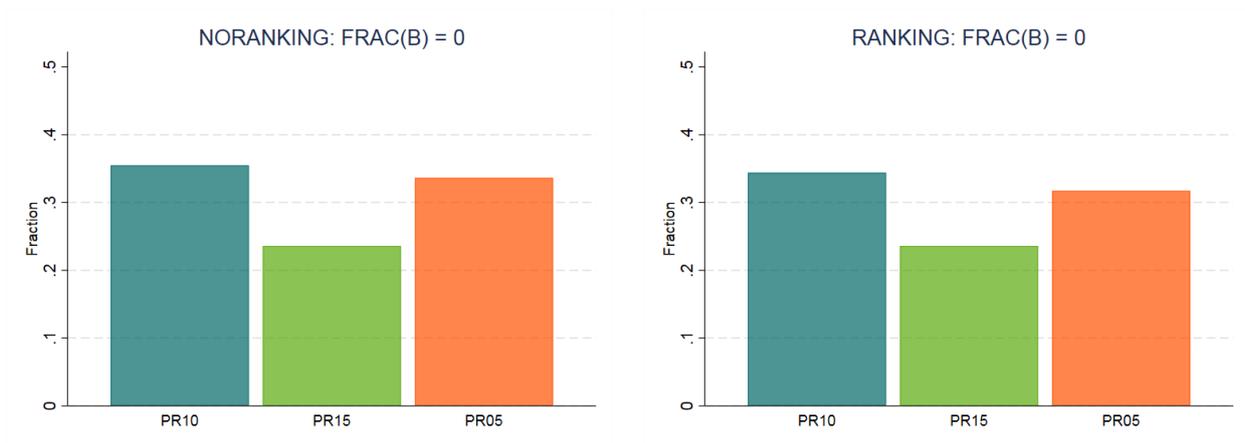


Figure B3: Fraction of subjects with zero relative effort in Domain B ( $FRAC(B) = 0$ ) in treatments with ranking (right) and without (left).

Table B3: Random-effects GLS regression results with TOTAL, the number of correct items in Domain A ( $A$ ) and in Domain B ( $B$ ) as dependent variables.  $A_{t-1}$  and  $B_{t-1}$ , respectively, convey the number of correct items in the previous period. DONATION indicates the subject's relevance of the donation. PR15 and PR05 are treatment dummies. RANKING indicates a dummy for the ranking treatments. Treatment PR10 serves as baseline. Variable  $Period^*$  ranges from 1 to 4 irrespective of random treatment allocation. Controls include gender, age, self-reported competitiveness, and self-reported risk tolerance. Robust standard errors are provided in parenthesis. \*  $p < 0.05$ , \*\*  $p < 0.01$ .

	<b>TOTAL</b>	<b>A</b>	<b>B</b>
$A_{t-1}$	0.075 (0.050)	0.369** (0.055)	-0.294** (0.048)
$B_{t-1}$	-0.028 (0.044)	-0.146** (0.029)	0.118* (0.045)
PR15	0.123 (0.159)	-0.045 (0.107)	0.168 (0.139)
PR05	0.217 (0.140)	0.009 (0.103)	0.208 (0.120)
RANKING	-0.092 (0.105)	-0.146 (0.091)	0.054 (0.108)
DONATION	-0.100 (0.067)	0.351** (0.058)	-0.450** (0.072)
$Period^*$	0.248** (0.060)	0.108* (0.044)	0.140* (0.053)
CONSTANT	3.099** (0.347)	0.921* (0.267)	2.177** (0.315)
$N$	858	858	858
$R^2$	0.045	0.408	0.325
$P > F$	0.000	0.000	0.000

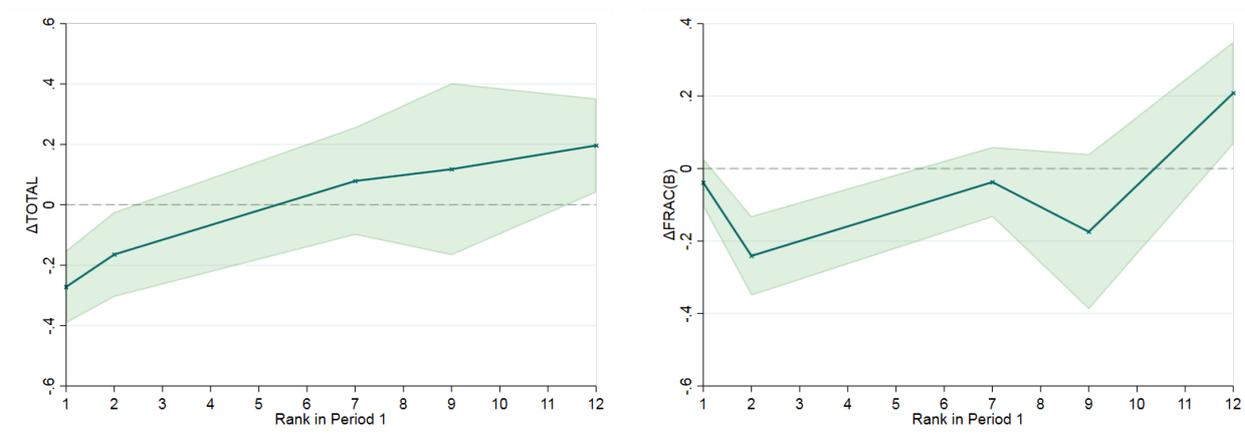


Figure B4: Percentage changes in the absolute number of solved items in total ( $\Delta\text{TOTAL}$ ; left) and in the fraction of solved items in the selfish activity of Domain B ( $\Delta\text{FRAC(B)}$ ; right) from period 1 to period 2 as a function of rank in period 1. Shaded areas represent 95% confidence intervals.

Table B4: OLS regression results with the following dependent variables: Percentage changes in the number of solved items ( $\Delta\text{TOTAL}$ ) and in the relative fraction of solved items in the self-ish activity of Domain B ( $\Delta\text{FRAC(B)}$ ) from period 1 to period 2. Variables  $\text{RANK}_{t=1}$  indicate dummies for an individual's rank in period 1 with  $\text{Rank} = 1$  serving as baseline.  $\Delta\text{NORANKING}$  indicates the changes as described in the respective treatments without ranking.  $\text{DONATION}$  stands for the self-reported relative of the donation.  $\text{PR15}$  and  $\text{PR05}$  indicate treatment dummies. Treatment  $\text{PR10}$  serves as baseline. Controls include gender, age, self-reported competitiveness, and self-reported risk tolerance. Robust standard errors are provided in parenthesis. \*  $p < 0.05$ , \*\*  $p < 0.01$ .

	$\Delta\text{TOTAL}$	$\Delta\text{FRAC(B)}$
$\text{RANK}_{t=1}$		
2	0.157 (0.109)	-0.258** (0.067)
7	0.402* (0.115)	-0.005 (0.065)
9	0.470* (0.159)	-0.086 (0.104)
12	0.615** (0.146)	0.397** (0.111)
$\Delta\text{NORANKING}$	-0.137 (0.073)	-0.008 (0.117)
$\text{PR15}$	0.083 (0.124)	0.164 (0.099)
$\text{PR05}$	0.115 (0.102)	0.133 (0.078)
$\text{DONATION}$	-0.054 (0.053)	-0.129** (0.037)
$\text{CONSTANT}$	-0.107 (0.214)	-0.173 (0.171)
$N$	286	286
$R^2$	0.089	0.132
$P > F$	< 0.000	0.002

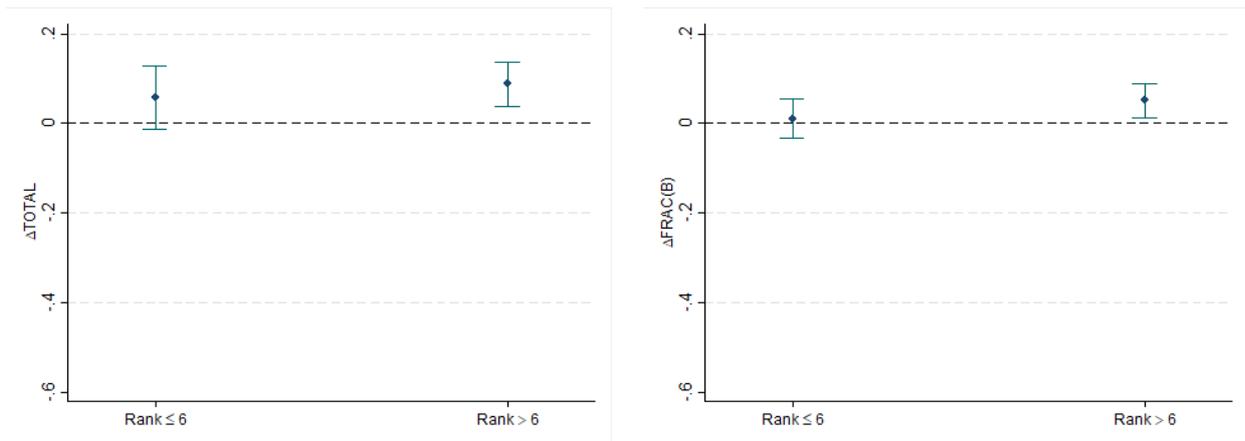


Figure B5: Percentage changes in the total number of solved items ( $\Delta_{TOTAL}$ ; left) and the relative fraction of solved items in the selfish activity of Domain B ( $\Delta_{FRAC(B)}$ ; right) from period 1 to period 2 in treatments without ranking. Outperformers (ranks  $\leq 6$ ) and underperformers (ranks  $> 6$ ) are defined by rankings in period 1 in treatment RANKING. Points indicate means and bars indicate standard errors.

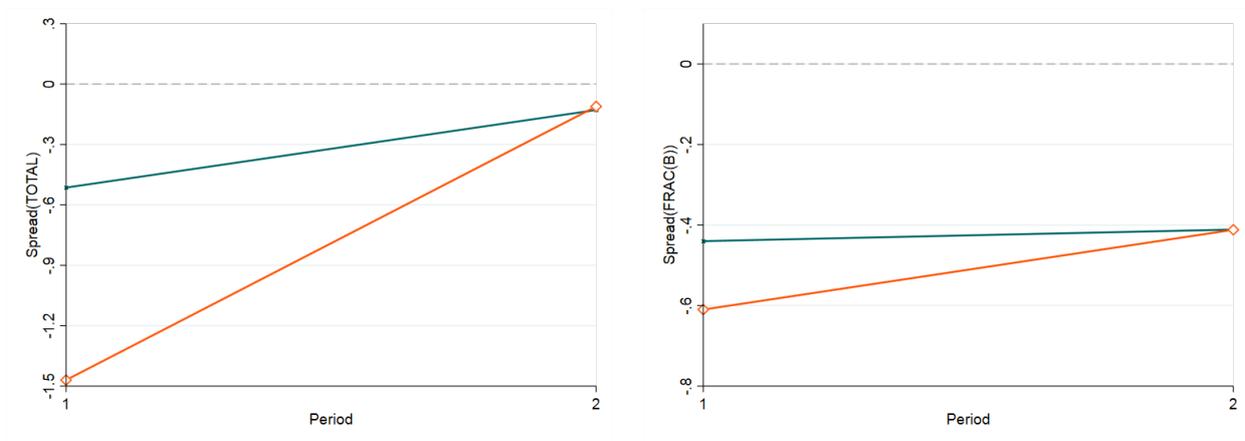


Figure B6: Behavior-spreads between out- (ranks  $\leq 6$ ) and underperformer (ranks  $> 6$ ) for treatment NORANKING (blue) and RANKING (orange). Calculation is based on mean TOTAL and FRAC(B), respectively, for out- and underperformer separately for periods 1 and 2. The behavior-spread is the difference between these means of underperformer and outperformer for each period separately. Note: out- and underperformers are defined by rankings in period 1 in treatment RANKING, also for treatment NORANKING.

## C Instructions of the experiment

Below screenshots of the experimental software are shown, except for the final payment screen. There are differences in the instructions of the two periods<sup>19</sup> and the feedback after each round between treatments with and without ranking. We have provided screenshots of each. In what follows, part 1 (2) corresponds to the treatment without (with) ranking. Note that both parts were randomized in the experiment to avoid ordering effects.

### Welcome

**Welcome to this experiment!**

We are happy you are participating in this experiment. Please do not use your mobile phone or tablet for this experiment. Moreover, we cannot guarantee full functionality, if you are using Internet Explorer or old versions of any web browser.

This online study adheres to the principles of economic experiments: participants are not deceived and all announced payments will really be made.

By participating in the experiment, you agree to our [terms of privacy](#)

Next

---

<sup>19</sup>Please note the difference in terminology between the main text (“period”) and the experimental software (“round”).

## General Overview

This experiment consists of 3 parts, which will be described in detail later. One of the first two parts is chosen with equal probability for your payout.

For participating in this experiment, you receive 10 Euros irrespective of any of the decisions you will make. You have to fully answer each part to receive a payment for the experiment.

In this experiment, you can choose to either earn donations to UNICEF for measles vaccine (one unit of vaccination costs 0.35 Euro) or to earn money for yourself. UNICEF describes the vaccination as follows (excerpt):

*Measles kills more than 160,000 children worldwide each year, yet it costs just 0.7 Euro to provide the two required doses to protect a child against the disease. Unicef is the world's largest provider of vaccines for global immunisation programmes, protecting millions of children every year from diseases like measles, tetanus and polio.*



(C) UNICEF



(C) UNICEF

*Please note:* These donations are actually made by us on your behalf. After every participant has finished this experiment, we will send you a confirmation of all received donations by UNICEF.

On a scale from 1 (not important at all) to 5 (very important): How important it is for you personally to make an effort to create donations to UNICEF for measles vaccine?

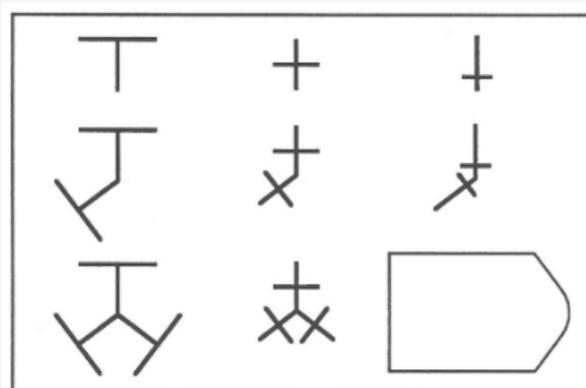
1  2  3  4  5

Next

## General Overview

While part 3 consists of a final questionnaire, the first two parts consist of **2 rounds of 2 minutes each**. In each round you have to solve intelligence-test-items. In case of a correct as well as a wrong answer, the next item is shown.

Each item comprises a pattern of diagrammatic puzzles with one piece missing. You are asked to choose the correct missing piece from a series of possible answers. The patterns in each item are presented in the form of a 3x3 matrix with the missing piece in the bottom right corner. An example of such a matrix is shown below.



Details about payouts and differences between the first two parts are described right before the decisions in the respective parts.

Next

## Instructions for Part 1

In each round, before the puzzle item is shown, you have to decide whether you want to solve the item for Domain A or Domain B: For each item you solve for **Domain A, 10 Euro will be donated to UNICEF** for measles vaccine. For each item you solve for **Domain B, you receive 5 Euro** as payout for yourself.

After each round you are told about the number of correct items you solved for Domain A and Domain B. In addition you will learn about your **rank** relative to **12 financial professionals**, who have participated in an experiment giving answers only for Domain B. **Thus, your ranking is based on your correct items in Domain B.** Each of the two rounds starts anew.

Next

## Instructions for Part 2

In each round, before the puzzle item is shown, you have to decide whether you want to solve the item for Domain A or Domain B: For each item you solve for **Domain A, 10 Euro will be donated to UNICEF** for measles vaccine. For each item you solve for **Domain B, you receive 5 Euro** as payout for yourself.

After each round you are told about the number of correct items you solved for Domain A and Domain B. Note that you will no longer be informed about the ranking shown in part 1. Each of the two rounds starts anew.

Next

# Domain Choice

Time left 1:58

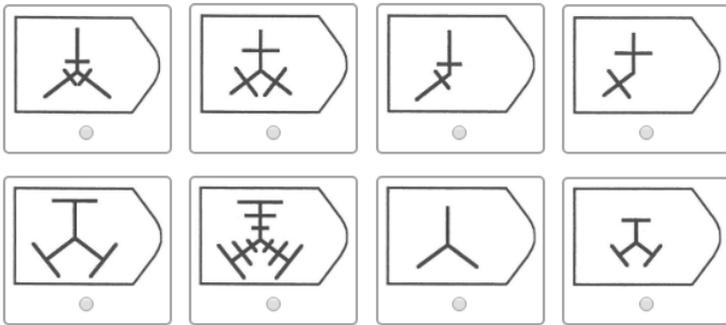
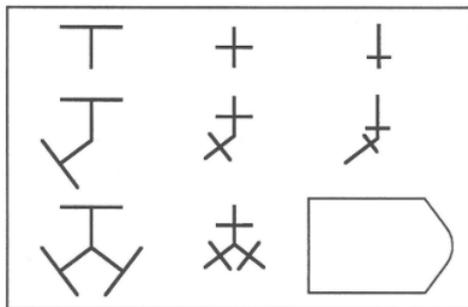
Please choose the domain for the next item:

- A: donation to UNICEF
- B: own payout/ ranking

Next

# Item Choice

Time left 1:26



Next

## Your Ranking

You correctly answered **0** item(s) in Domain A and **0** item(s) in Domain B.

Thus, your rank relative to the 12 financial professionals in Domain B is



To start with the second round, please continue.

Next

## Your Correct Items

You correctly answered **0** item(s) in Domain A and **0** item(s) in Domain B.

To start with the second round, please continue.

Next

## Final Questionnaire

Please answer the following short questions truthfully.

I enjoy working in situations involving competition with others (1 = strongly disagree; 5 = strongly agree).

1  2  3  4  5

It is important to me to perform better than others on a task (1 = strongly disagree; 5 = strongly agree).

1  2  3  4  5

I feel that winning is important in both work and games (1 = strongly disagree; 5 = strongly agree).

1  2  3  4  5

It annoys me when other people perform better than I do (1 = strongly disagree; 5 = strongly agree).

1  2  3  4  5

I try harder when I am in competition with other people (1 = strongly disagree; 5 = strongly agree).

1  2  3  4  5

How do you see yourself. Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? (0 = risk-averse; 10 = risk-prone)

0  1  2  3  4  5  6  7  8  9  10

Which job description fits best your current position?

Please indicate your gender:

Please indicate your age:

Next

## Your payout in this experiment

A random draw determined that Round 2 in Part 1 is relevant for your payout in the experiment. Therefore, your total earnings (including participation fee) in the experiment are **€10** and your donation is **€0**.

**Thank you very much for participating in the experiment!**

After every participant has finished this experiment, we will send you a confirmation of the donations by UNICEF. Please continue to facilitate your payment.

Next