

Stefan, Matthias; Huber, Jürgen; Kirchler, Michael; Sutter, Matthias; Walzl, Markus

Working Paper

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

Working Papers in Economics and Statistics, No. 2020-06

Provided in Cooperation with:

Institute of Public Finance, University of Innsbruck

Suggested Citation: Stefan, Matthias; Huber, Jürgen; Kirchler, Michael; Sutter, Matthias; Walzl, Markus (2020) : Monetary and Social Incentives in Multi-Tasking: The Ranking Substitution Effect, Working Papers in Economics and Statistics, No. 2020-06, University of Innsbruck, Research Platform Empirical and Experimental Economics (eeecon), Innsbruck

This Version is available at:

<https://hdl.handle.net/10419/238230>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

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

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

Working Papers in Economics and Statistics

2020-06



University of Innsbruck
Working Papers in Economics and Statistics

The series is jointly edited and published by

- Department of Banking and Finance
- Department of Economics
- Department of Public Finance
- Department of Statistics

Contact address of the editor:
research platform "Empirical and Experimental Economics"
University of Innsbruck
Universitaetsstrasse 15
A-6020 Innsbruck
Austria
Tel: + 43 512 507 71022
Fax: + 43 512 507 2970
E-mail: eeecon@uibk.ac.at

The most recent version of all working papers can be downloaded at
<https://www.uibk.ac.at/eeecon/wopec/>

For a list of recent papers see the backpages of this paper.

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

Matthias Stefan, Jürgen Huber, Michael Kirchler,
Matthias Sutter, and Markus Walzl[†]

October 15, 2020

Abstract

Rankings are intended as incentive tools on labor markets. Yet, when agents perform multiple tasks – not all of which can be ranked with respect to performance –, rankings might have unintended side-effects. Based on a dynamic model of multi-tasking, we present an experiment with financial professionals in which we identify hidden ranking costs when performance in one task is ranked while in another prosocial task it is not. We find that subjects lagging behind (leading) in the ranked task devote less (more) effort to the prosocial task. We discuss implications for optimal incentive schemes in organizations with multi-tasking.

JEL: C93, D02, D91

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

*Our experimental software, data files, and the analysis script file are stored on an OSF repository and will be made publicly 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.

[†]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. This is particularly important in cases of multi-tasking when a prosocial component is involved, as relative-performance evaluations of the non-prosocial task might potentially crowd out prosocial behavior and thus lead to efficiency losses for the company and society at large. Surprisingly, little scientific evidence exists on trade-off decisions between individual monetary and rank incentives on the one hand and prosocial behavior on the other hand.

In this paper, we narrow this research gap and investigate how monetary incentives and social comparison influence effort in a multi-tasking environment with a trade-off between one's own payment/rank and a prosocial activity. First, 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.¹ As a variation we introduce a ranking on the monetarily incentivized task, which informs the agent about her performance in this task relative to other agents. Depending on how the agent interprets the ranking, it can work as an additional incentive or disincentive for the monetarily incentivized task: If a good rank is desirable, it can increase 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 might lower total effort and the fraction of effort spent into the ranked task. Following the literature in social psychology (see e.g., Fishbach et al., 2009), one could expect the agent to balance (or highlight, i.e., focus on one of) the conflicting motives over time. For balancing, 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.

Second, we test the hypotheses derived in the model 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

¹Similar to Bénabou and Tirole (2016), one could interpret the first task as observable but not verifiable to the employer or the company.

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 (the 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 (high, medium and low 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 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 substitute relative effort spent for their own payment by putting more effort into the prosocial activity – i.e., the impact of a ranking is similar to a reduced piece-rate. In contrast, those lagging behind 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 act more prosocially, others focus more on the payoff-relevant task. 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 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 criticized as one of the potential drivers of excessive risk taking in the finance industry (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.

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 may have to trade off between high publication reputation

and contributing to the public good of service to the community and students.

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 Iriberri, 2010; Blanes-i-Vidal and Nossol, 2011; Tran and Zeckhauser, 2012) to differential effects depending on, for instance, expectations, current rank, details of the principal agent relationship and gender (Al-Ubaydli and List, 2015; Kuhnen and Tymula, 2012; Gill et al., 2019; Murad et al., 2019).² On a more general level, our paper also contributes to the literature on the relation between incentive schemes and performance measures (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, crucially depend on the specific set-up (Ariely et al., 2009; Falk and Szech, 2019). From this perspective, 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 the 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 existing studies mainly focus on effort in one domain, we extend the literature by studying a dynamic multi-tasking problem where subjects can distribute their effort between a selfish and a prosocial activity over time.³ The novel aspect of our design is that subjects can distribute effort between two activities (i.e., a feature of many real-world decisions) allowing us to separately analyze implications for total effort and for substitution effects between both activities.

2 The Model

A risk neutral agent decides how much effort to spend on two different domains (effort $a \geq 0$ in Domain A and effort $b \geq 0$ in Domain B) in each of $t = 1, \dots, T$ periods. In a given period t , efforts a_t and b_t generate convex costs $(a_t + b_t)^2/2$.⁴ The activity in Domain A is not verifiable

²Moreover, 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).

³For 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).

⁴Our 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.

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 .⁵ 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 , 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 is⁶

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

α measures how relevant the individual considers the donation. Φ is an increasing, weakly-concave and twice differentiable function of aggregate effort spent in Domain A. If Φ is linear, efforts in Domain A in periods before t have no impact on the marginal benefit of effort in Domain A in period t . If Φ 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). β 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. Ψ is a weakly-concave and twice differentiable function of $R(b_t)$. If Ψ is linear, efforts in Domain B in periods before t have no impact of the marginal benefit of effort in Domain B in period t . If Ψ 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, Ψ may be non-monotone in $R(b)$. E.g., if Ψ 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 $R(b) > R^*$.

⁵In an anonymized setting, (1) and (2) are straightforward implications of standard other-regarding preferences (see, e.g., Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000)) while (3) could be the result of self-image concerns (see, e.g., Bénabou and Tirole, 2006).

⁶For instance, in our experiment, described below, 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.

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 α , 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 α on a_t is stronger than the indirect (substitution) effect of α on b_t , the sum of the two efforts increases (as long as Ψ 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 Ψ is strictly concave and 0 in the absence of a ranking and if Ψ is linear, and (iv) $\frac{db_t^*/(a_t^*+b_t^*)}{d\alpha} < 0$.*

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 Φ 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 Φ is strictly concave and 0 if Φ is linear, and (iv) $\frac{db_t^*/(a_t^*+b_t^*)}{dy} > 0$.*

Observe that whenever Φ (Ψ) 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 Ψ 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 Ψ 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 Ψ 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 Φ is strictly concave and 0 if Φ is linear, and (iv) $\frac{db_t^*/(a_t^*+b_t^*)}{d\beta} > 0$.*

If, however, Ψ 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 $\sum_{\tau=1 \dots t} R(b_t) < R^*$, (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 Φ is strictly concave and 0 if Φ is linear, and (iv) $\frac{db_t^*/(a_t^*+b_t^*)}{d\beta} > 0$; for $\sum_{\tau=1 \dots t} R(b_t) > R^*$, (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 Φ is strictly concave and 0 if Φ 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, evidence shows that performance in IQ tests and related tasks are not only a measure of ability but also effort and do indeed respond to incentives (Borghans et al., 2013; Gneezy et al., 2019). Finally, there are no expected learning effects in APM.

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. In order to achieve a fair comparison, we applied the same two random sequences to 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; see also Kirchler et al. (2016) for a similar experimental design using measles vaccination as externality.) 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 an artefactual field experiment (Harrison and List, 2004), we implemented a within-between-subjects design (see Table 1 for details).

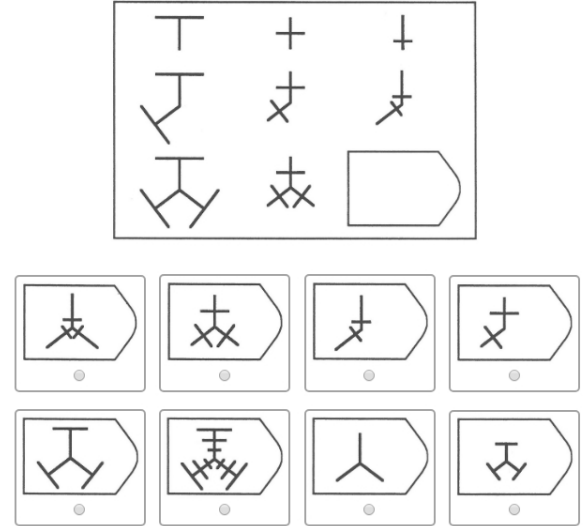


Figure 1: Left panel: Screenshot of the instruction screen outlining the explanation of the donation to measles vaccination to UNICEF (i.e., the prosocial activity). Right panel: Screenshot example of one APM item.

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 APM items 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. Given the possibilities of ties, participants could achieve ranks $\text{RANK} \in \{1, 2, 7, 9, 12\}$. 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 (i.e., including monetary incentives) except for two variations: First, there were only two periods, both without a ranking; second, subjects solved the items only for Domain B. This design choice avoids that participants might learn about social norms or the “relative return to effort” by observing the ranking. 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

Kirchler et al., 2018). We deliberately chose this design to implement the mildest form of social comparison: the ranking was anonymous, private, and had no monetary consequences. For this reason, it potentially addresses self-image concerns but not reputation or status, signalling, and learning from others.

As a 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 (which serves as baseline), 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: “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. 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).⁷ Finally, we also asked for participants’ age, gender and job

⁷The 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”).

description.⁸

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 ($SD = 8.61$). 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. Given our low gender ratio – which is typical for a financial professionals subject pool (see, e.g., Kirchler et al., 2018; Weitzel et al., 2020) – we cannot say much about gender effects (see, e.g., Murad et al., 2019). However, this is not the main question of our study, Rather, we primarily focus on professionals from an industry with a prominent ranking and competitive culture. 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 analysis 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.⁹

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.

⁸For an overview over participants’ self-reported characteristics see Table B1 in the Appendix.

⁹We chose the number of correct answers instead of the number of attempts 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 attempts noisy.

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) $\text{FRAC}(\text{B})$ 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 the 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 Ψ is monotonically increasing (i.e., a higher rank always induces a positive utility), Corollary 1 implies

Hypothesis 3 *(i) TOTAL is increasing in the presence of a ranking; (ii) $\text{FRAC}(\text{B})$ 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) $\text{FRAC}(\text{B})$ is increasing in the presence of a ranking. (2) If RANK_t is sufficiently good: (i) TOTAL is decreasing in the presence of a ranking; (ii) $\text{FRAC}(\text{B})$ 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 in this case.

4 Results of the experiment

Descriptive Results First, we show that the randomization of subjects into treatments resulted in the expected heterogeneity of types in each treatment: Pairwise tests of distributions (Kolmogorov-Smirnov tests) of self-reported relevance of donation (DONATION) do not reveal

significant differences between treatments ($p > 0.05$, $N \geq 182$ for each test). Distributions of normalized answers are depicted in Figure B2 in the Appendix.

Second, we provide a first overview over the two main variables of interest, TOTAL and FRAC(B) in Figure 2.¹⁰ We do not find major differences in the aggregate numbers for total effort (TOTAL) across treatments. In contrast, a higher 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 of the self-assessed relevance of the donation (DONATION) in Figure 3. 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. We find that the self-assessed relevance of the donation shows high explanatory power regarding the relative effort provided in Domain B. 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, invest all effort into one's own payment (Domain B). In fact, 39.86% (41.26%) of subjects spend all their effort in Domain B, i.e., $FRAC(B) = 1$, in treatments without (with) ranking. In contrast, 31.11% (30.07%) of subjects refuse to spend any effort in Domain B over both periods, i.e., $FRAC(B) = 0$, in treatments with (without) ranking (see also Figure B3 in the Appendix).

Result on Hypothesis 1 The total effort provided (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 (FRAC(B)) decreases with the self-assessed relevance of the donation.

To test Hypothesis 1, we run OLS regressions with TOTAL and FRAC(B), respectively, as dependent variables.¹¹ We add binary treatment indicators PR15 and PR05 as explanatory variables (i.e., treatment PR10 serves as baseline), a binary variable denoting treatments with ranking (RANKING), and a variable controlling for self-reported relevance of Domain A (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

¹⁰See 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 the analysis below, please also refer to panel regression results reported in Table B3 in the Appendix.

¹¹Please note that the variable 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.

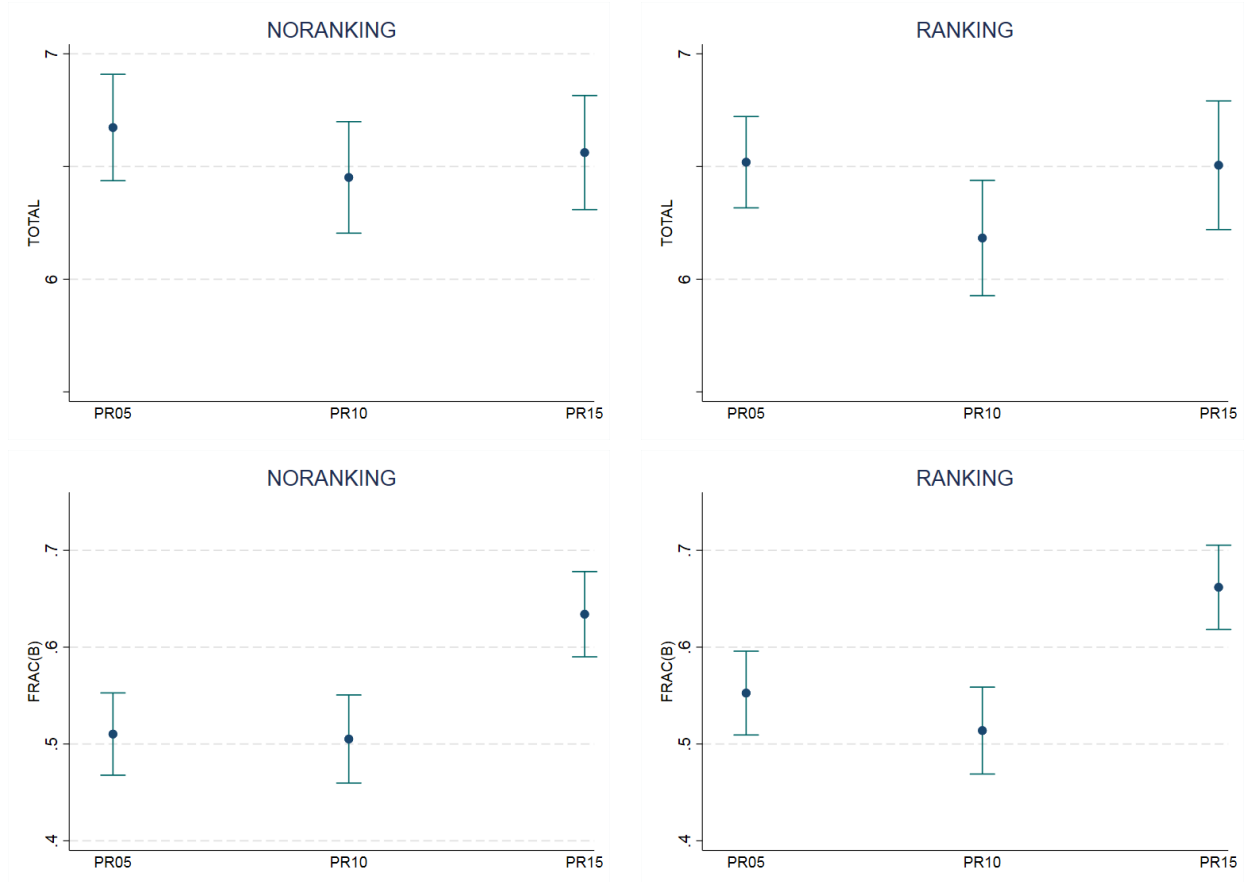


Figure 2: Upper panel: TOTAL denotes the total number of solved items in both Domains, A and B, across the two periods. Lower panel: FRAC(B) stands for the fraction of correct items in Domain B relative to the total number of correct items in both domains. PR05, PR10, and PR15 represent the different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

Table 2, TOTAL is independent of the self-assessed relevance of donation (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 (FRAC(B)) substantially decreases with donation relevance (DONATION).

Result on Hypothesis 2 Neither total effort provided (TOTAL), nor the effort provided for the selfish activity in Domain B (FRAC(B)) is significantly influenced by varying piece-rates.

As outlined in Table 2, we find that the piece-rates have neither impact on the total effort level (TOTAL), nor on relative effort in Domain B (FRAC(B)). Thus, we can reject Hypothesis 2 based on our results (see also Eckartz et al., 2012).

With regards to Hypotheses 1 and 2, we conclude that the individual preference for the

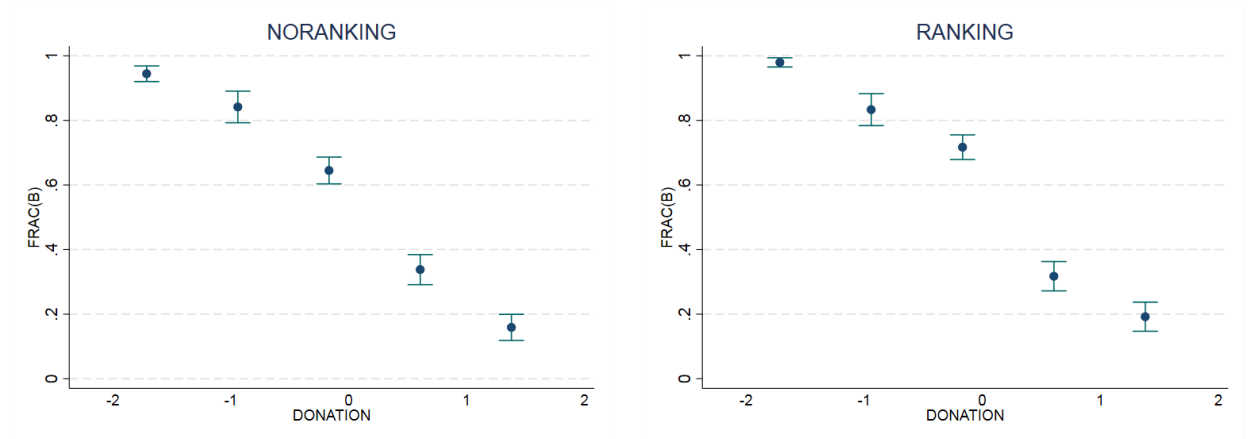


Figure 3: Mean $\text{FRAC}(B)$, i.e., the fraction of correct items in Domain B relative to the total number of correct items in both domains, as a function of the self-assessed relevance of the donation (DONATION). The right panel depicts observations from treatments with RANKING and the left panel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

donation has an impact on relative effort provided in Domain B, whereas the monetary incentives as piece-rates have no effect. Our findings are particularly noteworthy because of the absence of a substitution effect of monetary incentives.¹² While we cannot rule out effects with more substantial changes in piece-rates, we do not find evidence for a negative impact on prosocial behavior with moderately increasing monetary incentives.

Result on Hypothesis 3 The introduction of a ranking does not increase TOTAL and $\text{FRAC}(B)$ 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 (TOTAL) or in a higher relative effort in Domain B ($\text{FRAC}(B)$). In other words, we do not find an aggregate effect of rank incentives, which brings us to the next hypothesis.

Result on Hypothesis 4 The changes 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 $t = 1$ 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. We find no such effect for the total effort provided in both domains.

¹²See also Figure B3 in the Appendix depicting the fraction of subjects refusing to spend any effort in Domain B, i.e., $\text{FRAC}(B) = 0$.

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 relative to the total number of correctly solved items across both domains and both periods. DONATION indicates the subject’s relevance of the donation (in Domain A). PR05 and PR15 are dichotomous treatment indicators. Treatment PR10 serves as baseline. The binary variable RANKING indicates the ranking treatments. Controls include gender, age, self-reported competitiveness, and self-reported risk tolerance. Standard errors, clustered on the individual level, are provided in parenthesis. * $p < 0.05$, ** $p < 0.01$.

	TOTAL	FRAC(B)
PR05	0.324 (0.283)	0.008 (0.041)
PR15	0.145 (0.319)	0.078 (0.045)
RANKING	−0.161 (0.155)	0.027 (0.015)
DONATION	−0.069 (0.102)	−0.264** (0.015)
CONSTANT	7.718** (0.593)	0.715** (0.085)
N	572	572
R^2	0.048	0.455
$P > F$	0.021	< 0.001

To test Hypothesis 4 we examine percentage changes in total effort TOTAL and relative effort in Domain B FRAC(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 $\Delta\text{FRAC(B)} = \frac{\text{FRAC(B)}_{t=2} - \text{FRAC(B)}_{t=1}}{\text{FRAC(B)}_{t=1}}$.¹³ Importantly, since in our experiment, subjects face a random sequence of APM items, it may be that periods 1 and 2 vary in the difficulty of the items. Hence, for calculating ΔTOTAL , we adjust variable 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. ΔTOTAL 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 FRAC(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 TOTAL and, in turn, ΔTOTAL 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

¹³Please note, if $\text{TOTAL}_{t=1} = 0$ then $\Delta\text{TOTAL} = \text{TOTAL}_{t=2}$ and if $\text{FRAC(B)}_{t=1} = 0$, then $\Delta\text{FRAC(B)} = \text{FRAC(B)}_{t=2}$.

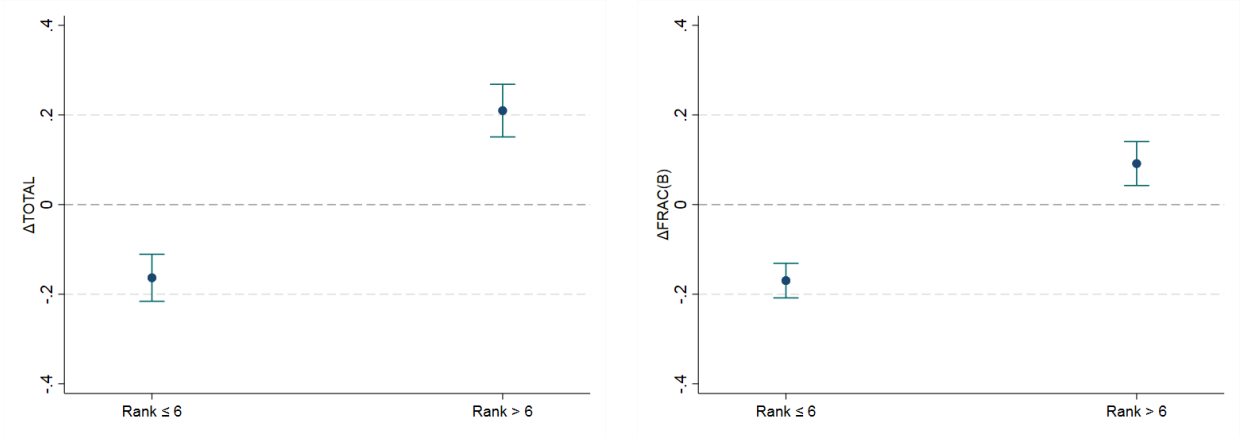


Figure 4: Percentage changes in the number of solved items in total (ΔTOTAL ; left) and in the fraction of solved items in Domain B ($\Delta\text{FRAC(B)}$; right) from period 1 to period 2 for outperformers (ranks ≤ 6) and underperformers (ranks > 6) in period 1. Points indicate means, bars standard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

ranking (ranks ≤ 6), on average, reduce total effort and decrease the fraction of solved items in Domain B. Instead, those lagging behind (ranks > 6), on average, do exactly the opposite. They increase total effort and switch more frequently to the selfish activity of Domain B. While this figure provides a first visual hint on the effect of a ranking, it cannot be ruled out that the observed dynamics in ΔTOTAL can be attributed to a regression to the mean-effect or subjects' balancing between the two domains and that the observed dynamics in $\Delta\text{FRAC(B)}$ would occur without an explicitly provided ranking, for instance due to an implicit comparison to expected outcomes, ranks, or non-ranking effects such as moral licensing, regret, or feedback.¹⁴

Thus, we test the ranking effect conjectured in Hypothesis 4 as a treatment comparison. To do so, we run OLS regressions with ΔTOTAL and $\Delta\text{FRAC(B)}$, respectively, serving as dependent variables (see Table 3). The main explanatory variables are the binary variable denoting treatments with ranking (RANKING) and the binary variable UNDERPERFORM, which takes the value 1 in case a subject's performance is below average and zero otherwise. In treatments with ranking, a subject's performance is below average with a rank of 7 or higher; in treatments without ranking the same level of underperformance is determined by less than 4 correct items for Domain B per period. Both specifications of underperformance are equivalent, i.e., a rank of 7 or lower is achieved by solving less than 4 items for Domain B.¹⁵ With this variable we

¹⁴For visual evidence that such dynamics are indeed relevant in treatments without ranking see Figure B5 in the Appendix.

¹⁵Please 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 a specification of OLS regression models with the individual's rank in period 1 ($\text{RANK}_{t=1}$) as explanatory variable instead of UNDERPERFORM are reported

can examine whether “out- and underperformers” react differently to the observed performance. The interaction between the treatment and underperformance variable is of particular interest, since it specifies the effect of showing a ranking in addition to the mere feedback in the baseline treatment – in other words, it provides evidence for a ranking effect controlled for a potential regression to the mean-effect or balancing of motives for both domains.¹⁶

Table 3: OLS regression results with the following dependent variables: percentage changes in the number of solved items (ΔTOTAL) and changes in the fraction of solved items in Domain B ($\Delta\text{FRAC(B)}$) from period 1 to period 2. The binary variable RANKING indicates the ranking treatments Variable UNDERPERFORM is a binary variable for underperforming subjects in period 1 ($\text{Rank} > 6$ or $B < 4$, respectively). PR05 and PR15 indicate treatments, i.e., PR10 serves as baseline. DONATION stands for the self-reported relevance of the donation. Controls include gender, age, self-reported competitiveness, and self-reported risk tolerance. Standard errors, clustered on the individual level, are provided in parenthesis. * $p < 0.05$, ** $p < 0.01$.

	ΔTOTAL	$\Delta\text{FRAC(B)}$
RANKING	−0.013 (0.074)	−0.082 (0.047)
UNDERPERFORM	0.366** (0.077)	0.164** (0.059)
RANKING*UNDERPERFORM	0.053 (0.110)	0.143* (0.072)
PR05	0.174** (0.064)	0.071 (0.057)
PR15	0.057 (0.071)	0.059 (0.066)
DONATION	−0.040 (0.029)	−0.032 (0.022)
CONSTANT	−0.079 (0.135)	−0.219 (0.130)
N	572	572
R^2	0.072	0.039
$P > F$	< 0.001	0.007

As shown in Table 3, we find that the period change in the absolute number of solved items not being different between treatments as the interaction term RANKING*UNDERPERFORM is small and not statistically significant. This could be an indicator of a regression to the mean or

in Table B5 in the Appendix. In these regression models Variable $\text{RANK}_{t=1}$ indicates a subject’s rank in period 1 or, for treatments without ranking, the corresponding number of correct items ($\text{Rank} = 1$ corresponds to $B \geq 5$, $\text{Rank} = 2$ to $B = 4$, $\text{Rank} = 7$ to $B = 3$, $\text{Rank} = 9$ to $B = 2$, and $\text{Rank} = 12$ to $B \leq 1$). Results remain qualitatively similar. See also Figure B4 in the Appendix.

¹⁶See Table B4 in the Appendix for regression models separated for each treatment.

balancing effect also existent in treatments without ranking (where the underperformer dummy is indeed significant). When focusing on the relative number of solved items ($\Delta\text{FRAC}(\text{B})$), we find a significant treatment effect of the ranking. In particular, those lagging behind in the ranking increase relative effort compared to those being ahead in the treatment without ranking. However, on average, this effect almost doubles in size in treatments with an explicit ranking, as indicated by the interaction term $\text{RANKING}*\text{UNDERPERFORM}$ (see also Table B4 in the Appendix). Thus, subjects in the ranking treatments react to the ranking relative to their peers presented in period 1: Those observing a below average relative performance (i.e., a bad rank) in period 1 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.

We, thus, 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 *relative* effort for the selfish domain. In contrast, those lagging behind, substitute between both domains by decreasing the number of solved items for the prosocial activity in Domain A, while increasing their relative effort in the selfish activity of Domain B. Interestingly, a similar, but significantly weaker effect also occurs without a ranking presented, which might be due to an implicit, subjective ranking or balancing motives such as moral licensing, regret, or received feedback. In contrast to substitution by domains, we do not find an effect of the observed ranking on the change in absolute number of solved item (ΔTOTAL).

Explanation of achieved ranks Additionally and in more exploratory terms, we examine which characteristics explain a subject’s rank. To do so, 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 , respectively), as well as a binary indicator for gender (FEMALE) and the age (AGE). For the regression on the individuals’ rank in period 2 in model 2 we also include $\text{RANK}_{t=1}$, i.e., the observed rank in period 1.

Results of the regression analyses are reported in Table 4. Please note that coefficients are reported as odds ratios. To begin with, 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 (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

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. DONATION stands for self-reported relevance of the donation, COMPETITIVENESS for self-reported competitiveness, and RISK for self-reported risk tolerance. Coefficients are reported as odds ratios. Robust standard errors are provided in parenthesis. * $p < 0.05$, ** $p < 0.01$.

	RANK_{t=1}	RANK_{t=2}
RANK _{t=1}		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
$P > F$	< 0.001	< 0.001

competitiveness by the externally provided ranking. Finally, we find that higher self-reported donation relevance (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), which, again, might be due to a crowding-out effect by the observed rank.

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 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 or the introduction of a ranking have a significant impact on *total* effort in our experiment.

However, participants *substitute* efforts spent in the two domains. The lower the relevance of the donation and the higher the piece-rate, 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, 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. Individuals lagging behind in the ranking tend to substitute effort in the pro-social task with effort in the task that determines earnings and the ranking position. For these individuals, a ranking has the same effect as a higher piece-rate. In contrast, individuals leading in the ranking substitute effort in the ranked activity with pro-social effort, i.e., these individuals react as if the donation has become more relevant.

Ranking Attenuation Effect: But if - in response to a ranking - an individual with a bad rank spends more effort in Domain B while an individual with a good rank spends less effort in Domain B, the difference in effort division across individuals tends to diminish because of the ranking. This is in stark contrast to the impact of rankings in single-task environments (without a prosocial dimension) where a ranking and tournament incentives are often used with the intention to increase effort differentials between agents who differ in productivity.¹⁷ 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

¹⁷As noted in Section 1, however, there is evidence of differential effects of rankings on effort, depending, among others, on expectations, current rank, and the principal agent relationship. See, e.g., Al-Ubaydli and List (2015); Blanes-i-Vidal and Nossol (2011); Tran and Zeckhauser (2012).

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 provision by underperforming agents, but introduces a disincentive for 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, potentially also 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 high 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 in 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 substituted relative effort spent for their own payment by putting more effort in the prosocial activity. In contrast, underperforming professionals substituted by spending more effort for their own payment and less for the prosocial activity.

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 affect 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.

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 Iriberri. 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.

- Murad, Zahra, Charitini Stavropoulou, Graham Cookson. 2019. Incentives and gender in a multi-task setting: An experimental study with real-effort tasks. *PLOS ONE* **14**(3) 1–18.
- 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. *eeecon Working Paper 2020-19* .
- Weitzel, Utz, Christoph Huber, Jürgen Huber, Michael Kirchler, Florian Lindner, Julia Rose. 2020. Bubbles and financial professionals. *Review of Financial Studies* **forthcoming**.

Appendix

A Proofs

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

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

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 Φ and Ψ (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 $()$.

Proof of Proposition 1:

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

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

Proof of Proposition 2:

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

(iii) With (i) and (ii), $\frac{d(a+b)}{dy} = -\frac{\alpha \frac{\partial^2 \Phi}{\partial a^2}}{() } > 0$ if Φ is strictly concave (and $= 0$ if Φ 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 $\text{FRAC}(\text{B})$, 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. DONATION indicates the subject’s self-reported relevance of the donation. PR05 and PR15 are dichotomous treatment indicators, with PR10 serving as baseline.. RANKING indicates ranking treatments. Controls include gender, age, self-reported competitiveness, and self-reported risk tolerance. Standard errors, clustered on the individual level, are provided in parenthesis. * $p < 0.05$, ** $p < 0.01$.

	FRAC(B)
PR05	−0.007 (0.142)
PR15	0.256 (0.168)
RANKING	0.101 (0.053)
DONATION	−0.883** (0.074)
CONSTANT	0.738* (0.316)
N	572
$PseudoR^2$	0.293
$P > F$	< 0.001

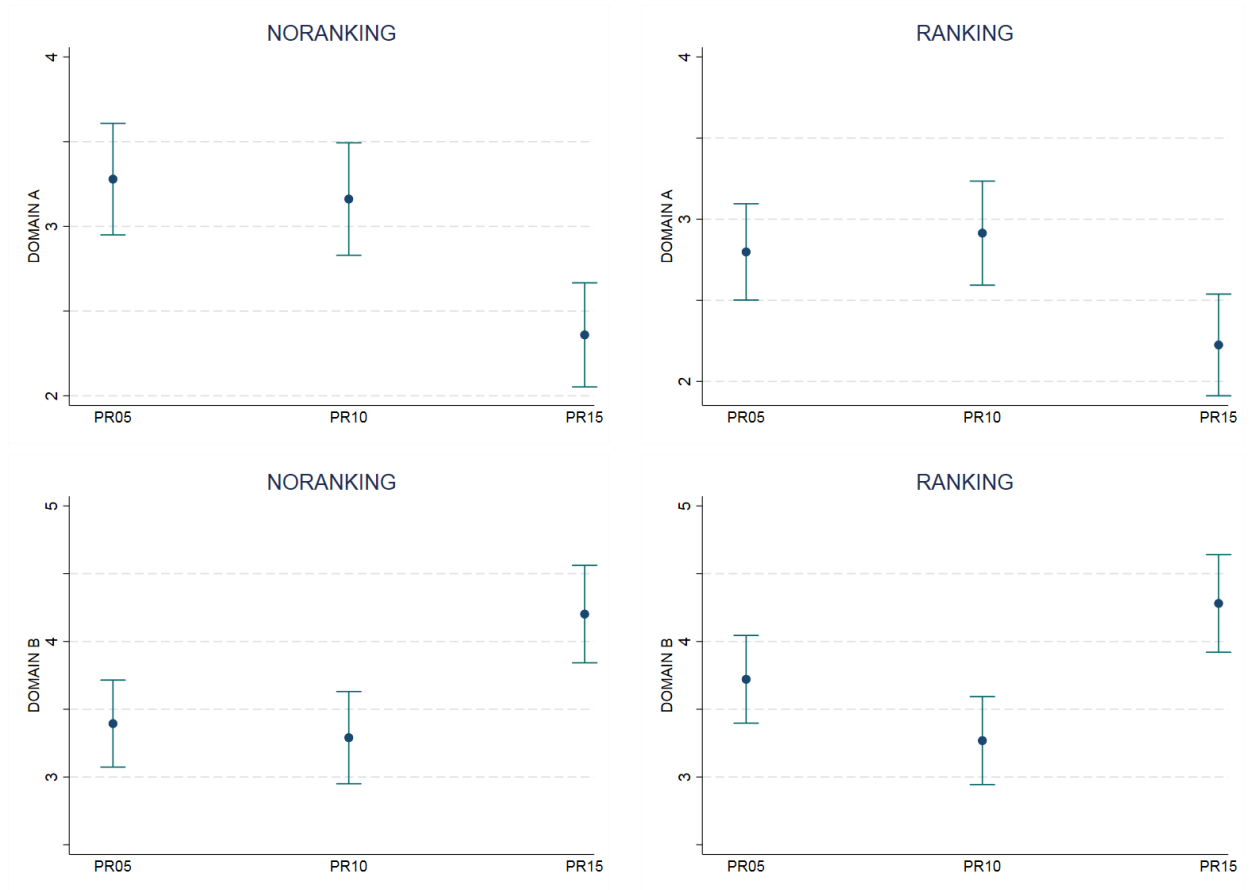


Figure B1: PR05, PR10, and PR15 represent the different treatments with RANKING (right panel) and without (left panel). Points indicate means, bars standard errors.

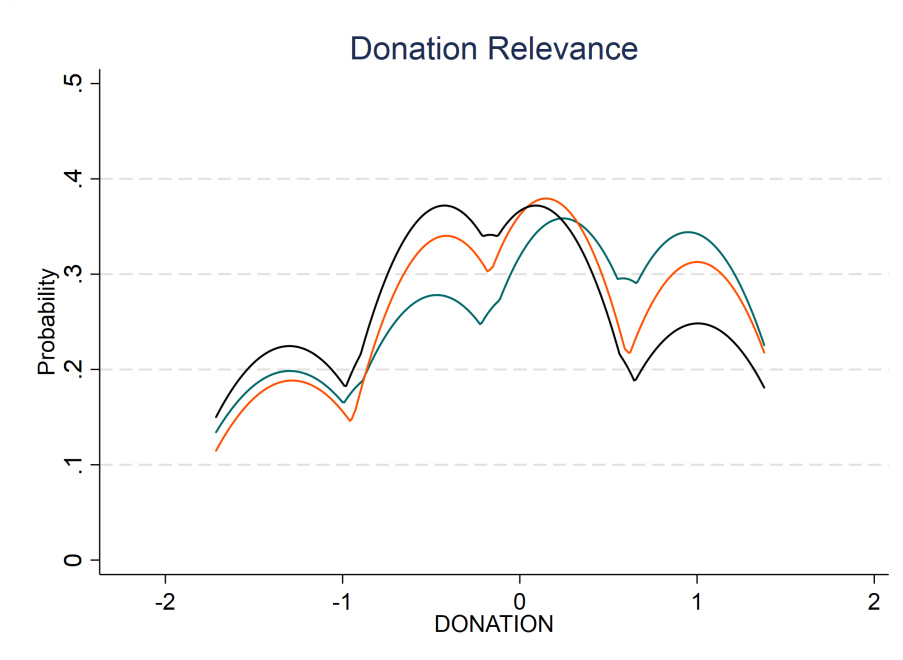


Figure B2: Distribution of the normalized variable for self-reported relevance of the donation (DONATION) for each treatment. Treatment PR05 is depicted by black lines, PR10 by blue lines, and PR15 by orange lines.

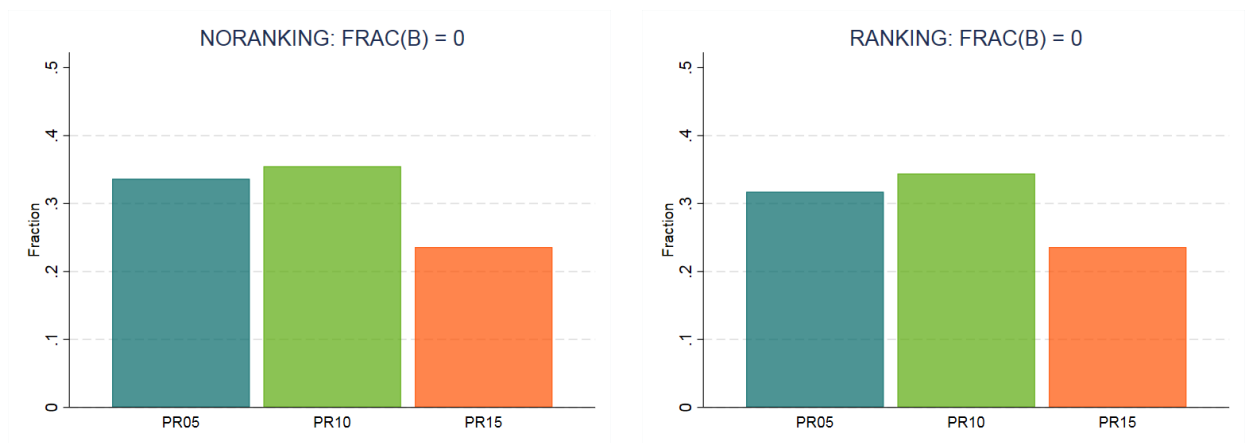


Figure B3: Fraction of subjects with zero relative effort in Domain B ($\text{FRAC}(B) = 0$) in treatments with ranking (right) and without (left).

Table B3: Random-effects GLS regression results with TOTAL, the number of correct items in Domain A (A), and the number of correct items 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. PR05 and PR15 indicate treatments, with PR10 serving as baseline. RANKING indicates ranking treatments. Variable $Period^*$ ranges from 1 to 4 irrespective of random treatment allocation. Controls include gender, age, self-reported competitiveness, and self-reported risk tolerance. Standard errors, clustered on the individual level, are provided in parenthesis. * $p < 0.05$, ** $p < 0.01$.

	TOTAL	A	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)
PR05	0.217 (0.140)	0.009 (0.103)	0.208 (0.120)
PR15	0.123 (0.159)	-0.045 (0.107)	0.168 (0.139)
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
$P > F$	< 0.001	< 0.001	< 0.001

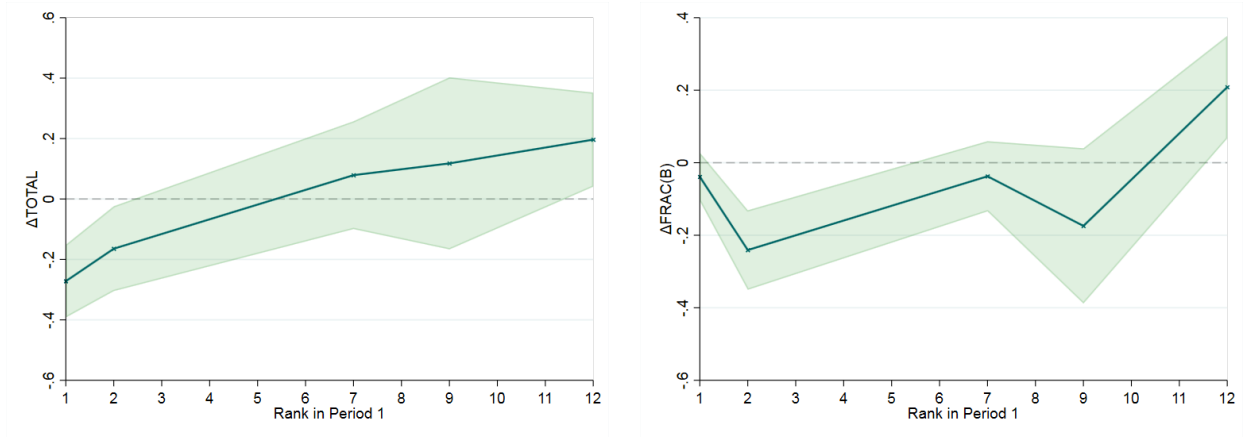


Figure B4: Percentage changes in the absolute number of solved items in total (ΔTOTAL ; left) and in the fraction of solved items in Domain B ($\Delta\text{FRAC(B)}$; right) from period 1 to period 2 as a function of rank in period 1. Shaded areas represent 95% confidence intervals.

Table B4: OLS regression results with the following dependent variables: percentage changes in the number of solved items (ΔTOTAL) and changes in the fraction of solved items in Domain B ($\Delta\text{FRAC(B)}$) from period 1 to period 2. Variable UNDERPERFORM is a binary variable for underperforming subjects in period 1 ($\text{Rank} > 6$ or $B < 4$, respectively). PR05 and PR15 indicate treatments, with PR10 serving as baseline. DONATION indicates the self-reported relevance of the donation. Controls include gender, age, self-reported competitiveness, and self-reported risk tolerance. Robust standard errors are provided in parenthesis. * $p < 0.05$, ** $p < 0.01$.

	NORANKING		RANKING	
	ΔTOTAL	$\Delta\text{FRAC(B)}$	ΔTOTAL	$\Delta\text{FRAC(B)}$
UNDERPERFORM	0.359** (0.080)	0.152* (0.064)	0.407** (0.094)	0.321** (0.073)
PR05	0.251** (0.094)	0.027 (0.076)	0.096 (0.103)	0.117 (0.080)
PR15	0.025 (0.087)	0.005 (0.086)	0.087 (0.125)	0.113 (0.100)
DONATION	-0.057 (0.041)	-0.015 (0.031)	-0.021 (0.044)	-0.050 (0.031)
CONSTANT	-0.091 (0.216)	-0.221 (0.167)	-0.067 (0.230)	-0.300 (0.190)
N	286	286	286	286
R^2	0.103	0.030	0.064	0.054
$P > F$	< 0.001	0.114	< 0.001	0.010

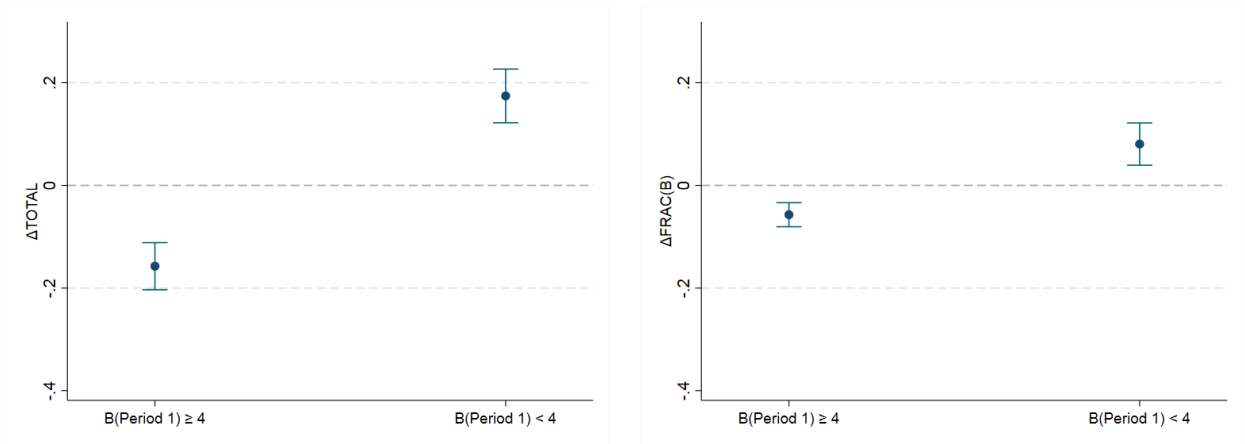


Figure B5: Percentage changes in the total number of solved items (Δ_{TOTAL} ; left) and the relative fraction of solved items in Domain B ($\Delta_{\text{FRAC(B)}}$; right) from period 1 to period 2 in treatments without ranking. Outperformers (underperformers) are defined by the number of solved items in Domain B in period 1 as ≥ 4 (< 4). Points indicate means, bars standard errors.

Table B5: OLS regression results with the following dependent variables: percentage changes in the number of solved items (ΔTOTAL) and changes in the fraction of solved items in Domain B ($\Delta\text{FRAC(B)}$) from period 1 to period 2. Variable $\text{RANK}_{t=1}$ indicates a subject's rank in period 1 or, for treatments without ranking the corresponding number of correct items ($\text{Rank} = 1$ corresponds to $B \geq 5$, $\text{Rank} = 2$ to $B = 4$, $\text{Rank} = 7$ to $B = 3$, $\text{Rank} = 9$ to $B = 2$, and $\text{Rank} = 12$ to $B \leq 1$). PR05 and PR15 indicate treatments, with PR10 serving as baseline. DONATION indicates the self-reported relevance of the donation. Controls include gender, age, self-reported competitiveness, and self-reported risk tolerance. Robust standard errors are provided in parenthesis. * $p < 0.05$, ** $p < 0.01$.

	NORANKING		RANKING	
	ΔTOTAL	$\Delta\text{FRAC(B)}$	ΔTOTAL	$\Delta\text{FRAC(B)}$
$\text{RANK}_{t=1}$				
2	0.263** (0.086)	-0.029 (0.064)	0.133 (0.109)	-0.258** (0.066)
7	0.487** (0.098)	-0.097 (0.077)	0.384** (0.118)	-0.005 (0.063)
9	0.365** (0.136)	0.019 (0.135)	0.455** (0.161)	-0.087 (0.103)
12	0.599** (0.117)	0.342** (0.102)	0.593** (0.144)	0.396** (0.106)
PR05	0.227* (0.094)	-0.002 (0.073)	0.082 (0.105)	0.133 (0.078)
PR15	0.028 (0.089)	0.029 (0.084)	0.085 (0.125)	0.164 (0.098)
DONATION	-0.078 (0.051)	-0.088* (0.041)	-0.053 (0.053)	-0.129** (0.037)
CONSTANT	-0.240 (0.223)	-0.145 (0.163)	-0.115 (0.222)	-0.172 (0.178)
N	286	286	286	286
R^2	0.122	0.100	0.074	0.132
$P > F$	< 0.001	0.003	< 0.001	0.001

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 periods¹⁸ and the feedback after each period 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

¹⁸Please 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?

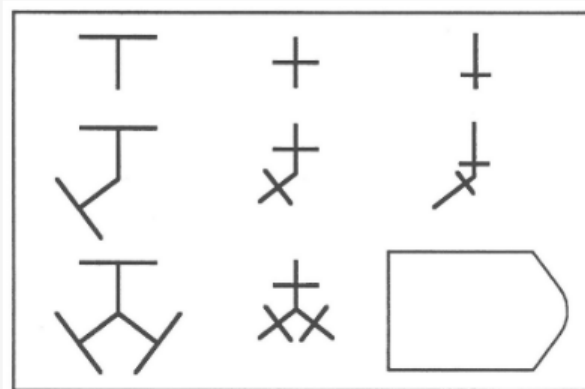
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

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**

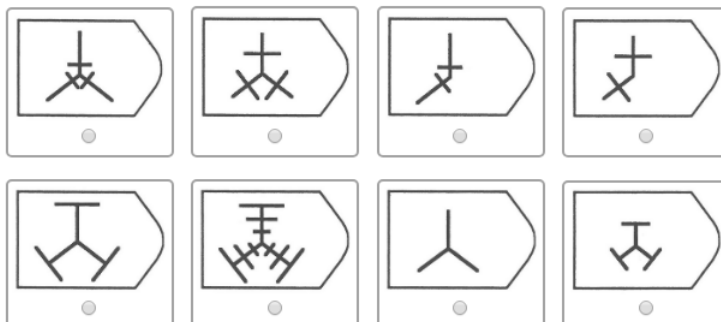
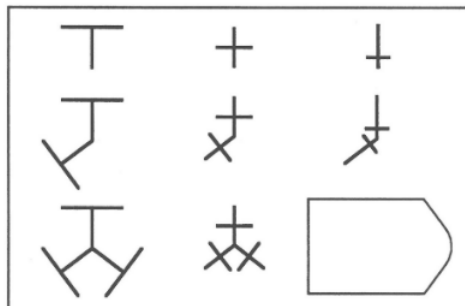
Please choose the domain for the next item:

- ☐ 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

University of Innsbruck - Working Papers in Economics and Statistics

Recent Papers can be accessed on the following webpage:

<https://www.uibk.ac.at/eeecon/wopec/>

- 2020-06 **Matthias Stefan, Jürgen Huber, Michael Kirchler, Matthias Sutter, Markus Walzl:** Monetary and Social Incentives in Multi-Tasking: The Ranking Substitution Effect
- 2020-05 **Michael Razen, Jürgen Huber, Laura Hueber, Michael Kirchler, Matthias Stefan:** Financial Literacy, Economic Preferences, and Adolescents' Field Behavior
- 2020-04 **Christian König-Kersting, Johannes Lohse, Anna Louisa Merkel:** Active and Passive Risk-Taking
- 2020-03 **Christoph Huber, Jürgen Huber:** Bad bankers no more? Truth-telling and (dis)honesty in the finance industry
- 2020-02 **Dietmar Fehr, Daniel Müller, Marcel Preuss:** Social Mobility Perceptions and Inequality Acceptance
- 2020-01 **Loukas Balafoutas, Rudolf Kerschbamer:** Credence goods in the literature: What the past fifteen years have taught us about fraud, incentives, and the role of institutions
- 2019-21 **Felix Holzmeister, Martin Holmen, Michael Kirchler, Matthias Stefan, Erik Wengström:** Delegated Decision-Making in Finance
- 2019-20 **Julia Rose, Michael Kirchler, Stefan Palan:** Status and Reputation Nudging
- 2019-19 **Felix Holzmeister, Matthias Stefan:** The risk elicitation puzzle revisited: Across-methods (in)consistency?
- 2019-18 **Katharina Momsen, Markus Ohndorf:** Information Avoidance, Selective Exposure, and Fake(?) News-A Green Market Experiment
- 2019-17 **Stjepan Srhoj, Bruno Skrinjaric, Sonja Radas, Janette Walde:** Closing the Finance Gap by Nudging: Impact Assessment of Public Grants for Women Entrepreneurs
- 2019-16 **Adam Farago, Martin Holmen, Felix Holzmeister, Michael Kirchler, Michael Razen:** Cognitive Skills and Economic Preferences in the Fund Industry
- 2019-15 **Christopher Kah, Daniel Neururer:** Generiert der stationäre Buchhandel positive Nachfrageeffekte und verhilft dadurch dem Kulturgut Buch bei seiner Verbreitung? - Ein natürliches Experiment
- 2019-14 **Stjepan Srhoj, Michael Lapinski, Janette Walde:** Size matters? Impact evaluation of business development grants on SME performance

- 2019-13 **Andrea M. Leiter, Engelbert Theurl:** Determinants of prepaid systems of healthcare financing - A worldwide country-level perspective
- 2019-12 **Michael Razen, Michael Kirchler, Utz Weitzel:** Domain-Specific Risk-Taking Among Finance Professionals
- 2019-11 **Jonathan Hall, Rudolf Kerschbamer, Daniel Neururer, Eric Skoog:** Uncovering sophisticated discrimination with the help of credence goods markups - evidence from a natural field experiment
- 2019-10 **Daniela Glätzle-Rützler, Philipp Lergetporer, Matthias Sutter:** Collective intertemporal decisions and heterogeneity in groups
- 2019-09 **Morten Hedegaard, Rudolf Kerschbamer, Daniel Müller, Jean-Robert Tyran:** Distributional Preferences Explain Individual Behavior Across Games and Time
- 2019-08 **Daniel Müller, Sander Renes:** Fairness Views and Political Preferences - Evidence from a representative sample
- 2019-07 **Florian Lindner, Michael Kirchler, Stephanie Rosenkranz, Utze Weitzel:** Social Status and Risk-Taking in Investment Decisions
- 2019-06 **Christoph Huber, Julia Rose:** Individual attitudes and market dynamics towards imprecision
- 2019-05 **Felix Holzmeister, Jürgen Huber, Michael Kirchler, Florian Lindner, Utz Weitzel, Stefan Zeisberger:** What Drives Risk Perception? A Global Survey with Financial Professionals and Lay People
- 2019-04 **David M. McEvoy, Tobias Haller, Esther Blanco:** The Role of Non-Binding Pledges in Social Dilemmas with Mitigation and Adaptation
- 2019-03 **Katharina Momsen, Markus Ohndorf:** When do people exploit moral wiggle room? An experimental analysis in a market setup
- 2019-02 **Rudolf Kerschbamer, Daniel Neururer, Matthias Sutter:** Credence goods markets and the informational value of new media: A natural field experiment
- 2019-01 **Martin Geiger, Eric Mayer, Johann Scharler:** Inequality and the Business Cycle: Evidence from U.S. survey data

University of Innsbruck

Working Papers in Economics and Statistics

2020-06

Matthias Stefan, Jürgen Huber, Michael Kirchler, Matthias Sutter, and Markus Walzl

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

Abstract

Rankings are intended as incentive tools on labor markets. Yet, when agents perform multiple tasks - not all of which can be ranked with respect to performance -, rankings might have unintended side-effects. Based on a dynamic model of multi-tasking, we present an experiment with financial professionals in which we identify hidden ranking costs when performance in one task is ranked while in another prosocial task it is not. We find that subjects lagging behind (leading) in the ranked task devote less (more) effort to the prosocial task. We discuss implications for optimal incentive schemes in organizations with multitasking.

ISSN 1993-4378 (Print)

ISSN 1993-6885 (Online)